

Validierung des Desired Level of Confidence

Inaugural-Dissertation
zur Erlangung der Doktorwürde
der
Philosophischen Fakultät
der
Rheinischen Friedrich-Wilhelms-Universität
zu Bonn

vorgelegt von

Marc Jekel

aus
Bonn

Bonn, 2012

Gedruckt mit der Genehmigung der Philosophischen Fakultät
der Rheinischen Friedrich-Wilhelms-Universität Bonn

Zusammensetzung der Prüfungskommission:

Prof. Dr. Gerhard Blickle
(Vorsitzender)

Prof. Dr. Arndt Bröder
(Betreuer und Gutachter)

Prof. Dr. Rainer Banse
(Gutachter)

Prof. Dr. Georg Rudinger
(weiteres prüfungsberechtigtes Mitglied)

Tag der mündlichen Prüfung: 30.09.2011

Für die Betreuung und Unterstützung danke ich Herrn Prof. Dr. Arndt Bröder. Für die Abnahme der Prüfung danke ich Herrn Prof. Dr. Rainer Banse, Herrn Prof. Dr. Gerhard Blickle und Herrn Prof. Dr. Georg Rudinger. Für die reibungslose Koordination meiner Disputation danke ich sehr herzlich Frau Roswitha Mombauer. Für die Unterstützung und einen wertvollen Ideenaustausch danke ich Frau Dipl.-Psych. Christine Platzer und Frau Dr. Julia Schütz. Für Korrekturvorschläge und gute Diskussionen danke ich Frau Pia Jüngermann-Fröhlich. Besonderer Dank geht an meine vielen Probanden, ohne die diese Arbeit nicht möglich gewesen wäre. Zuletzt danke ich den *intuitive experts* für die Unterstützung (nicht nur aber besonders) in den letzten Wochen vor der Disputation.

(auch) Meiner Familie

Abstract

In der vorliegenden Arbeit wird in acht Studien das *Desired Level of Confidence* (DLC) von Hausmann und Läge (2008) validiert. Das DLC ist ein individuelles Maß des Abbruchs der Informationssuche im Entscheidungsprozess: Erst wenn die Validität einer Information über dem individuellen *Anspruchsniveau an gewünschter Urteilssicherheit* (Hausmann-Thürig, 2004) liegt, wird die Suche nach weiteren Informationen abgebrochen und eine Entscheidung getroffen. Das DLC kann in einer Entscheidungsaufgabe - der Pferdewette - gemessen werden. In Studie 1 wurde untersucht, ob das DLC in der Messaufgabe reliabel erfasst werden kann. Dazu wurde die Vorhersagekraft des DLC für das Entscheidungsverhalten in einem unabhängigen Set von Aufgaben bestimmt. Weiterhin wurde untersucht, inwieweit das DLC von irrelevanten Merkmalen der Messaufgabe abhängig ist. Es zeigt sich, dass das DLC Modell Entscheidungsverhalten gut vorhersagen kann und dass eine Veränderung der Merkmale der Messaufgabe keinen Einfluss auf die interindividuellen Unterschiede im DLC hat; die Korrelation zwischen dem DLC der Pferdewette und dem DLC der für die Studie entwickelten Messaufgabe - dem Kriminalfall - liegt bei einem Pearson's $r = .79$. In Studie 2 wurde untersucht, wie zeitlich stabil das DLC ist und inwieweit es sich damit beim DLC um ein Persönlichkeitsmerkmal handelt. Weiterhin wurde getestet, wie das DLC mit Prozessmaßen einer Entscheidung und mit der Art der Integration von Informationen zusammenhängt. Es zeigt sich, dass die Messungen des DLC über einen Zeitraum von einer Woche mit einem Pearson's $r = .61$ zeitlich stabil ist, dass Probanden mit einem höheren DLC tendenziell mehr Information in einer Entscheidungsaufgabe - dem Börsenspiel - aufdecken (Pearson's $r = .42$) und dass das DLC nicht mit der Art der Integration der Informationen zusammenhängt. In den folgenden vier Studien wurde untersucht, ob das DLC mit situationalen Bedingungen interagiert. In Studie 3 und 4 wurde untersucht, ob ein Priming eines One-Reason Entscheidungsverhaltens (ORDM) - das ist ein Entscheidungsverhalten, in dem *wenige* Informationen berücksichtigt werden - vs. More-Reason Entscheidungsverhalten (MRDM) - das ist ein Entscheidungsverhalten, in dem *viele* Informationen berücksichtigt werden - zu einem niedrigen vs. hohen DLC führt. Es zeigt sich, dass weder ein konzeptuelles Priming noch ein Mindset Priming einen Einfluss auf die Höhe des DLC hat. In Studie 5 und 6 wurde untersucht, ob Probanden, die ein hohes vs. niedriges situationales *Need for Cognitive*

Closure (Kruglanski & Webster, 1996) induziert bekommen, ein geringes vs. hohes DLC haben. Es zeigt sich, dass weder die Attraktivität der Folgeaufgabe noch ein aversiver Ton in der Erhebungssituation auf die Höhe des DLC wirken. In den abschließenden Studien wurde untersucht, ob das DLC mit Persönlichkeitsvariablen zusammenhängt. In Studie 7 und 8 wurde untersucht, ob das DLC mit Antworttendenzen in einer Signalentdeckungsaufgabe zusammenhängt. Es zeigt sich, dass Probanden mit einem hohen DLC weder in einer Kategorisierungsaufgabe noch in einer Rekognitionsaufgabe ein signifikant konservativeres Entscheidungskriterium wählen. Es zeigt sich über die Studien hinweg kein stabiler signifikanter Zusammenhang zwischen dem DLC und dem *Bedürfnis nach kognitiver Geschlossenheit (16-NCCS)* (Schlink & Walther, 2007), dem regulatorischen Fokus (Werth & Förster, 2007), der *Achievement Motives Scale (AMS-R)* (Lang & Fries, 2006) und - bis auf einen signifikant positiven Zusammenhang mit der Offenheit für Erfahrung (Pearson's $r = .27$) - mit den *Big Five* der Persönlichkeit (Borkenau & Ostendorf, 1993). Statistische Probleme, wie die geringe Teststärke in den Studien, konzeptuelle Probleme und methodische Probleme, wie die verwendete Messmethode zur Bestimmung des DLC, werden abschließend diskutiert.

Inhaltsverzeichnis

| | |
|--|-----------|
| 1. Theorie | 13 |
| 1.1. Begrenzte Rationalität | 17 |
| 1.2. Entscheidungsheuristiken und Umwelten | 25 |
| 1.2.1. Parameter eines Cues | 25 |
| 1.2.2. ORDM: Entscheidungen basierend auf einem Cue | 27 |
| 1.2.3. MRDM: Entscheidungen basierend auf mehreren Cues | 30 |
| 1.2.4. Überblick Studien zur normativen Güte | 31 |
| 1.2.5. Überblick Studien zur deskriptiven Güte | 38 |
| 1.2.5.1. <i>Take The Best</i> Heuristik | 42 |
| 1.2.5.2. Rekognitions-Heuristik | 46 |
| 1.2.6. Abschließende Bewertung | 50 |
| 1.3. Unimodelle einer Entscheidung | 52 |
| 1.3.1. Parallele Informationsverarbeitung | 52 |
| 1.3.2. Sequentielle Informationsverarbeitung | 56 |
| 1.3.2.1. Aufgabe Desired Level of Confidence . . | 60 |
| 1.3.2.2. Messmodell Desired Level of Confidence | 62 |
| 1.3.2.3. Befunde Desired Level of Confidence . . | 65 |
| 1.4. Überblick empirischer Teil der Arbeit | 66 |
| 2. Reliabilität und Validität | 69 |
| 2.1. Reliabilität des DLC | 69 |
| 2.1.1. Methoden | 70 |
| 2.1.1.1. Stichprobe | 74 |
| 2.1.2. Hypothesen | 74 |
| 2.1.3. Ergebnisse | 74 |
| 2.1.4. Diskussion | 77 |
| 2.2. Validität des DLC | 79 |
| 2.2.1. Methoden | 80 |
| 2.2.1.1. Stichprobe | 81 |
| 2.2.1.2. Prozessmaße einer Entscheidung | 81 |
| 2.2.1.3. Bayesianische Strategieklassifikation . . | 83 |

| | | |
|-----------|---|------------|
| 2.2.2. | Hypothesen | 83 |
| 2.2.3. | Ergebnisse | 84 |
| 2.2.4. | Diskussion | 90 |
| 3. | Situative Einflussfaktoren | 93 |
| 3.1. | Priming | 93 |
| 3.1.1. | Konzeptuelles Priming: Sätze bilden | 93 |
| 3.1.1.1. | Methoden | 95 |
| 3.1.1.2. | Hypothesen | 97 |
| 3.1.1.3. | Ergebnisse | 97 |
| 3.1.2. | Mindset Priming: Geschichten erfinden | 100 |
| 3.1.2.1. | Methoden | 101 |
| 3.1.2.2. | Hypothesen | 104 |
| 3.1.2.3. | Ergebnisse | 104 |
| 3.1.3. | Diskussion | 106 |
| 3.2. | Need for Cognitive Closure | 108 |
| 3.2.1. | Aufgabenattraktivität: Videos anschauen | 109 |
| 3.2.1.1. | Methoden | 110 |
| 3.2.1.2. | Hypothesen | 112 |
| 3.2.1.3. | Ergebnisse | 112 |
| 3.2.2. | Erhebungssituation: Lärm ertragen | 113 |
| 3.2.2.1. | Methoden | 113 |
| 3.2.2.2. | Hypothesen | 117 |
| 3.2.2.3. | Ergebnisse | 117 |
| 3.2.3. | Diskussion | 118 |
| 4. | Korrelation mit Persönlichkeitsvariablen | 121 |
| 4.1. | Signalentdeckung: Antworttendenzen | 121 |
| 4.1.1. | Zuordnungsaufgabe: Objekte klassifizieren | 124 |
| 4.1.1.1. | Methoden | 125 |
| 4.1.1.2. | Hypothesen | 129 |
| 4.1.1.3. | Ergebnisse | 129 |
| 4.1.2. | Rekognition: Gesichter wiedererkennen | 130 |
| 4.1.2.1. | Methoden | 131 |
| 4.1.2.2. | Hypothesen | 132 |
| 4.1.2.3. | Ergebnisse | 132 |
| 4.1.3. | Diskussion | 133 |
| 4.2. | Andere Persönlichkeitsvariablen | 135 |

| | |
|---|------------|
| 5. Diskussion | 141 |
| 5.1. Statistische Gründe | 141 |
| 5.2. Konzeptuelle Gründe | 142 |
| 5.3. Methodische Gründe | 143 |
| 5.4. Fazit | 145 |
| Literatur | 147 |
| Abbildungsverzeichnis | 162 |
| Tabellenverzeichnis | 165 |
| A. Formeln | 167 |
| A.1. Bayesianische Strategieklassifikation | 167 |
| A.2. Validitäten erster Cue der DLC Mess- und Testphase . . | 171 |
| B. Instruktionen Studien | 175 |
| B.1. Instruktion der Pferdewette | 175 |
| B.2. Studie 1: Instruktion des Börsenspiels | 176 |
| B.3. Studie 2: Instruktion des Kriminalfalls | 177 |
| B.4. Studie 5: Instruktion Need For Cognitive Closure | 178 |
| B.5. Studie 7: Instruktion Kategorisierungsaufgabe | 179 |
| C. Aufgaben Studien | 183 |
| C.1. Studie 2: Entscheidungsszenarien | 183 |
| C.2. Studie 3: Sätze Konzeptuelles Priming | 188 |
| C.3. Studie 3: Szenarien Konzeptuelles Priming | 190 |
| C.4. Studie 4: Verwendete Begriffe Mindset Priming | 191 |
| C.5. Studie 8: Oberfläche Pferdewette | 193 |

1. Theorie

Welche Stadt hat mehr Einwohner: Stuttgart oder Nürnberg? Falls einem Entscheider¹ die Anzahl der Einwohner beider Städte nicht bekannt und nicht zugänglich ist, muss er, wenn er nicht raten möchte, Informationen, die für die Entscheidung relevant und verfügbar sind, heranziehen (Gigerenzer & Goldstein, 1996, 1999).²

Diese Informationen können internal im Gedächtnis (Bröder, 2007; Bröder & Schiffer, 2003b, 2006b) und/oder external in der Umwelt (Payne, Bettman & Johnson, 1988, 1993) “gesucht” werden. In der Städtevergleichsaufgabe können das beispielhaft die Informationen sein, ob eine der Städte eine Universität hat, ob eine der Städte eine Fußballmannschaft in der ersten Bundesliga hat oder ob eine der Städte Landeshauptstadt ist (vgl. Gigerenzer & Goldstein, 1996, Tabelle 1, S. 655). Nachdem der Entscheider die Informationssuche abbricht oder abbrechen muss, z. B. aus Zeitgründen (Payne et al., 1988), können die vorhandenen Informationen genutzt werden, um zu einer Entscheidung zu gelangen.

Wenn der Entscheider z. B. meint, dass Landeshauptstädte mehr Einwohner haben, wird er sich, wenn er rational handelt, aufgrund der Information, dass Stuttgart eine Landeshauptstadt ist und Nürnberg nicht, für Stuttgart als die Stadt mit mehr Einwohnern entscheiden.

An dieser Entscheidungsaufgabe lassen sich Grundbegriffe und Grundkonzepte aus der Entscheidungspsychologie erläutern. Es gibt Entscheidungsoptionen wie in diesem Beispiel zwei Städte. Es gibt eine an einem Außenkriterium validierbare korrekte Antwort³, d. h., entweder Stuttgart oder Nürnberg hat mehr Einwohner oder beide Städte haben

¹Definition: Ein Entscheider ist eine Person, die Entscheidungen trifft.

²Zwar gibt es Studien, die auf die Relevanz von Geschlechtsunterschieden bei Entscheidungen (z. B. Huang & Wang, 2010; Powell & Ansic, 1997) hindeuten, Geschlechtsunterschiede sind aber nicht Thema der Arbeit. “Entscheider” ist somit im Folgenden geschlechtsneutral zu verstehen.

³Dieser Aufgabentyp ist von einer Präferenztaufgabe abzugrenzen, bei der die Konsequenzen der Optionen bekannt sind und die Wahrscheinlichkeit des Eintretens und die Präferenz für die Konsequenzen die Entscheidung beeinflussen. Eine Präferenzentscheidung kann nicht logisch falsch sein, sie kann nur suboptimal sein, wenn man normative Standards als Vergleich heranzieht (Kahneman &

1. Theorie

gleich viele Einwohner. Die korrekte Antwort ist unbekannt, also muss eine Entscheidung unter Unsicherheit getroffen werden. Als Entscheidungshilfe können Informationen konsultiert werden, in dem obigen Städtebeispiel sind das drei mögliche binäre Informationen (Universität vorhanden versus nicht vorhanden etc.).

Die Anzahl der Optionen und das Messniveau der Informationen sind freie Parameter, d. h., es sind auch Entscheidungsaufgaben mit mehr als zwei Optionen - wie z. B. dem Vergleich zwischen drei und mehr Städten - basierend auf metrischen Informationen - wie z. B. dem prozentualen Anteil an Grünflächen in Städten - möglich.

Die Informationen, die man auch Hinweisreize oder Cues nennt (Bröder, 2009; Brunswik, 1955; Gigerenzer & Goldstein, 1996), deuten nicht mit vollkommener Sicherheit auf die korrekte Antwort hin. Es ist nicht immer der Fall, dass eine Landeshauptstadt im Vergleich zu einer deutschen Stadt, die nicht Landeshauptstadt ist, mehr Einwohner hat, wie es z. B. bei dem Vergleich zwischen Frankfurt und Stuttgart der Fall ist (vgl. Gigerenzer & Goldstein, 1996, Appendix, S. 668). Da die Cues nur mit einer gewissen Wahrscheinlichkeit auf die korrekte Antwort deuten, handelt es sich um eine probabilistische Inferenzaufgabe. Die Höhe dieser Wahrscheinlichkeit ist ein Gütekriterium eines Cues und wird die ökologische Validität eines Cues genannt (Bröder, 2009; Gigerenzer & Goldstein, 1996).

Um einen Cue gewinnbringend einsetzen zu können, muss die Validität des Cues über der Ratewahrscheinlichkeit ($p = .5$ bei zwei Optionen) liegen, um überzufällig häufig korrekte Entscheidungen treffen zu können. Weiterhin muss der Zusammenhang der Richtung der Cue Ausprägung mit dem Entscheidungskriterium bekannt sein, d. h., welche Cue Ausprägung auf eine größere Stadt deutet, um nicht überzufällig häufig schlechte Entscheidungen zu treffen.

Ein Cue ist jedoch nur dann diagnostisch für eine Vergleichsaufgabe, wenn er zwischen den Entscheidungsoptionen diskriminiert. Wenn der Cue für alle Entscheidungsoptionen die gleiche Ausprägung hat, kann aufgrund des Cues keine Entscheidung zwischen den Optionen getroffen werden. Stuttgart und Nürnberg sind beides Universitätsstädte. Es lässt sich also anhand des Universität-Cues keine Entscheidung zwischen den Städten treffen. Die Rate, mit der ein Cue zwischen Optionen

Tversky, 1979, 2000). Eine Präferenzentscheidung kann jedoch im Vergleich zu anderen Präferenzentscheidungen inkonsistent sein, was die Rationalität eines Entscheiders in Frage stellt (siehe z. B. framing Effekte; Tversky & Kahneman, 1981).

diskriminiert, nennt man Diskriminanzrate (Bröder, 2009; Gigerenzer & Goldstein, 1996, 1999); siehe Kapitel 1.2.1 (S. 25) für eine mathematische Definition der ökologischen Validität und der Diskriminanzrate eines Cues.

Die Entscheidungsfindung ist ein Prozess, d. h., im einfachsten Fall werden Cues gesucht, die Suche nach Cues wird abgebrochen und basierend auf den konsultierten Cues wird eine Entscheidung getroffen (Gigerenzer & Goldstein, 1996, 1999; Gigerenzer, Todd & The ABC Research Group, 1999).

Wie die Entscheidung getroffen wird, d. h., ob und wie die Cues zu einer Entscheidung integriert werden, ist gemäß dem Forschungsprogramm der adaptiven Werkzeugkiste (Gigerenzer et al., 1999) abhängig von der Entscheidungsstrategie, die ein Entscheider für die Entscheidungsaufgabe wählt. Denkbar ist eine Entscheidung, die nur auf dem Landeshauptstadt-Cue basiert; denkbar ist aber auch eine Entscheidung, die auf der Verrechnung einzelner oder aller Cues basiert. In Kapitel 1.2.2 (S. 27) und Kapitel 1.2.3 (S. 30) werden einzelne Strategien genauer vorgestellt.

Bei der Beurteilung der Nützlichkeit von Entscheidungsstrategien sind normative und deskriptive Aspekte zu trennen. Wie gut eine Entscheidungsstrategie im Vergleich zu anderen Entscheidungsstrategien Entscheidungsaufgaben korrekt lösen kann, ist eine normative Frage. Die Bestimmung der optimalen Entscheidung wird mit steigender Anzahl der Cues computational nicht lösbar, so dass es unwahrscheinlich ist, dass menschliche Entscheider aufgrund von mentalen und situationalen Beschränkungen die rational - im Sinne des Kohärenzkriteriums (d. h. im Einklang mit der normativ korrekten Verrechnung von Cues nach der Wahrscheinlichkeitstheorie, siehe auch S. 20 für weitere Erläuterungen) - beste Entscheidungsstrategie anwenden können (Bröder, 2000c; Gigerenzer & Todd, 1999; Martignon & Hoffrage, 2002; Oaksford & Chater, 1993; Simon, 1956). In Kapitel 1.1 (S. 17) wird das Problem der begrenzten Rationalität eines menschlichen Entscheiders und dem daraus hervorgegangenen Forschungsprogramm der adaptiven Werkzeugkiste (Gigerenzer, Hoffrage & Kleinböling, 1991; Gigerenzer et al., 1999) beschrieben. Als Kriterium für einen Strategievergleich kann z. B. die Anzahl der korrekten Entscheidungen in Entscheidungsaufgaben wie der oben beschriebenen Städteaufgabe dienen (Czerlinski, Gigerenzer & Goldstein, 1999; Gigerenzer & Brighton, 2009); In Kapitel 1.2.4 (S. 31) wird ein Überblick über Studien zur normativen Güte von Entscheidungsstrategien gegeben.

1. Theorie

Der Nachweis über die Akkuratheit einer Strategie in einer Entscheidungsaufgabe und damit über die Plausibilität der Anwendung einer Strategie sagt noch nichts darüber aus, ob und in welchem Ausmaß Menschen eine Strategie auch tatsächlich anwenden (Bröder & Schiffer, 2003b, S. 278). Wie gut eine Entscheidungsstrategie menschliche Entscheidungen vorhersagen und Entscheidungsprozesse beschreiben kann und unter welchen Bedingungen eine bestimmte Strategie gegenüber einer anderen Strategie bevorzugt wird, ist eine empirische Frage (Bröder, 2000b, 2000c; Bröder & Newell, 2008; Newell & Shanks, 2007). Die mathematische Spezifikation des Entscheidungsprozesses einer Entscheidungsstrategie lässt empirische Vorhersagen zu und macht es möglich, die Passung der Vorhersagen einer Entscheidungsstrategie zu beobachtetem Entscheidungsverhalten zu bestimmen (Bröder & Schiffer, 2003a; Glöckner, 2009; Jekel, Nicklisch & Glöckner, 2010). Es ist dann möglich, Einflussfaktoren auf die Strategiewahl zu untersuchen. In Kapitel 1.2.5 (S. 38) wird ein Überblick zu den Studien zur deskriptiven Güte von Entscheidungsstrategien gegeben.

Neben der Idee der adaptiven Werkzeugkiste, dass menschliche Entscheider über ein Repertoire an Entscheidungsstrategien verfügen, finden sich in der Literatur als Gegenentwurf Unimodelle einer Entscheidung (Busemeyer & Townsend, 1993; Dougherty, Gettys & Ogden, 1999; Glöckner & Betsch, 2008a; Hausmann & Läge, 2008; Juslin & Persson, 2002; Lee & Cummins, 2004). Diese Modelle haben gemeinsam, dass sie menschliches Entscheiden über *ein* Modell abbilden; siehe Kapitel 1.3 (S. 52) für einen Überblick über Unimodelle menschlichen Entscheidens.

Das Entscheidungsmodell zum Desired Level of Confidence (Hausmann & Läge, 2008) - übersetzt „Anspruchsniveau an gewünschter Urteilsicherheit“ (Hausmann-Thürig, 2004) und im Folgenden mit DLC abgekürzt -, das zu der Klasse der Unimodelle einer Entscheidung gehört, ist Thema dieser Arbeit. In diesem Modell wird angenommen, dass Entscheider eine individuell messbare gewünschte Schwelle an Urteilsicherheit anstreben, bevor sie die Suche nach weiteren Cues in einer Entscheidungsaufgabe abbrechen und eine Entscheidung treffen (Chaiken, Liberman & Eagly, 1989; Hausmann & Läge, 2008). In Kapitel 1.3.2.1 (S. 60) wird die Aufgabe zur Messung des DLC, in Kapitel 1.3.2.2 (S. 62) das Messmodell des DLC und in Kapitel 1.3.2.3 (S. 65) die empirischen Befunde zum DLC Modell beschrieben.

Da es zu dem Modell nur eine publizierte Studie gibt (Hausmann & Läge, 2008), bedarf es einer empirischen Validierung des Modells.

Es ist abzuklären, ob das individuell messbare DLC ein psychologisch bedeutsames Konstrukt abbildet und in welcher Relation das DLC-Modell zu den Entscheidungsstrategien der adaptiven Werkzeugkiste steht (Konstrukvalidität), ob es sich bei dem DLC um eine Persönlichkeitseigenschaft handelt (Stabilität) und/oder ob das DLC mit situationalen Manipulationen interagiert und mit konstruktverwandten Persönlichkeitsvariablen korreliert ist (experimentelle Studien). Im abschließenden Kapitel 1.4 (S. 66) des Theorieteils sind die Ziele der Arbeit und der Aufbau des empirischen Teils der Arbeit genauer beschrieben.

1.1. Begrenzte Rationalität

Die Städtevergleichsaufgabe (Gigerenzer & Goldstein, 1996, 1999), die in der Einführung beschrieben wird, ist bei zwei möglichen Alternativen und drei Cues übersichtlich. Ein Entscheider, dessen Entscheidung auf einem einzigen Cue basiert, wendet eine Entscheidungsstrategie aus der Klasse des *One-Reason Decision Making* (Hausmann & Läge, 2008) - im Folgenden mit ORDM abgekürzt - an. Bei drei Cues ergeben sich 3!, d. h. also $3 \times 2 \times 1 = 6$, mögliche Reihenfolgen, die Cues sequentiell zu konsultieren.

Erhöht sich die Anzahl der Cues, steigen die möglichen Suchreihenfolgen jedoch an: Nach Martignon (1999, S. 152) ist das Auffinden der optimalen Suchreihenfolge „NP-complete“, das heißt nicht in polynomialer Zeit durchführbar und somit für eine große Zahl von Cues computationally nicht mehr zu bewältigen. In der Originalstudie zur Städtevergleichsaufgabe (Gigerenzer & Goldstein, 1996, 1999) lagen neun Cues vor, so dass es also 9! mögliche Suchreihenfolgen gibt. Zur Bestimmung einer optimalen Suchreihenfolge müssen also über das gesamte Set von Städten - in der Studie waren das 83 Städte und damit 3,403 mögliche Vergleiche⁴, 362,880 mögliche Suchreihenfolgen - auf ihre Performanz hin, d. h. die relative Anzahl der korrekt gelösten Städtevergleiche, evaluiert werden.

Martignon und Schmitt (1999) konnten zeigen, dass es keine ökonomischere Methode als die exhaustive Suche über alle möglichen Cue-Suchreihenfolgen zum Auffinden der optimalen Suchreihenfolge gibt

⁴In der Arbeit wird die amerikanische Zeichensetzung für Zahlen verwendet, d. h., dass der Punkt die Dezimalstellen und das Komma die Tausenderstellen einer Zahl kenntlich macht.

1. Theorie

(vgl. auch mit Martignon & Hoffrage, 2002, Theorem 2, S. 39). Dass eine optimale Ordnung in einem Set von Vergleichen von Objekten nicht eine optimale Ordnung in einem unabhängigen Set von Objekten garantiert, soll hier erwähnt, aber noch nicht thematisiert werden; siehe dazu die Anmerkungen zur Kreuzvalidierung auf S. 33 in Kapitel 1.2.4.

Wenn alle Cues zugänglich sind, es also Zeit- oder Kostengründe nicht verhindern, dass mehr als ein Cue konsultiert und für eine Entscheidung verwendet wird, kann eine Entscheidungsstrategie aus der Klasse des *More-Reason Decision Making* (Hausmann & Läge, 2008) - im Folgenden mit MRDM abgekürzt - angewendet werden. Die Wahrscheinlichkeit, dass Stadt A größer ist als Stadt B bei einem gegebenem Cue-Muster, kann über die Anwendung des Bayes-Theorems (Bayes, 1763; Gigerenzer, Czerlinski & Martignon, 1999) berechnet werden; siehe Lee und Cummins (2004, S. 345) für eine mathematische Darstellung für die beschriebene Städtevergleichsaufgabe. Der Aufwand der Berechnung erhöht sich bei steigender Anzahl von Cues, was sich bei Cues, die untereinander korreliert sind, besonders drastisch zeigt (Gigerenzer & Goldstein, 1996; Lee & Cummins, 2004).

Eine menschliche Entscheidung wird jedoch unter Bedingungen getroffen, die durch die Umwelt und die kognitive Ausstattung eines Entscheiders moderiert wird. Die Anforderungen der Umwelt, wie z. B. eine Entscheidung unter Zeitdruck fällen zu müssen (Payne et al., 1988), lassen es nicht zu, die optimale Suchreihenfolge von Cues zu ermitteln. Die kognitiven Fähigkeiten, wie z. B. die begrenzte Arbeitsgedächtniskapazität (Baddeley, Eysenck & Anderson, 2009; Miller, 1956), lassen es nicht zu, die Option mit der höchsten Wahrscheinlichkeit für eine korrekte Antwort nach Bayes zu berechnen. Ein menschlicher Entscheider kann somit den normativ besten Algorithmus zur Optimierung einer Entscheidung in der Regel nicht einsetzen. Er kann nur begrenzt rational entscheiden (Simon, 1956).

Eine Möglichkeit für eine begrenzt rationale Entscheidung ist die Anwendung von Heuristiken (Gilovich, Griffin & Kahneman, 2002; Tversky & Kahneman, 1974, 1983). Heuristiken sind kognitive Daumenregeln, die sich der Performanz von optimalen Algorithmen bei gleichzeitig vermindertem Rechenaufwand annähern. Heuristiken können jedoch zu systematischen Fehlern in einer Entscheidung führen.

Ein Anwender der Repräsentativitäts-Heuristik schließt von der Passung eines Ereignisses auf ein mentales Modell, wie z. B. einem Prototyp oder Schema, auf die Wahrscheinlichkeit für das Auftretens des Ereignisses (Tversky & Kahneman, 1983, S. 295). Ein (A) sonniger und

heißer Tag ist für einen prototypischen Sommertag repräsentativer als (B) ein regnerischer und kalter Tag, was sich in Wahrscheinlichkeitsurteilen für das Auftreten des Ereignisses A und B widerspiegeln sollte (Tversky & Kahneman, 1983).

Die Repräsentativitäts-Heuristik führt jedoch zu systematischen Fehlern wie dem Konjunktionsfehler (Tversky & Kahneman, 1983). Die Wahrscheinlichkeit, dass ein Ereignis A auftritt, kann niemals kleiner als die Wahrscheinlichkeit sein, dass Ereignis A und ein weiteres Ereignis B gemeinsam auftreten. Es gilt also $p(A \cap B) \leq p(A)$ (Kahneman & Tversky, 1996; Tversky & Kahneman, 1983). Die Wahrscheinlichkeit, dass Linda⁵ Bankkauffrau ist, ist geringer als die oder gleich der Wahrscheinlichkeit, dass Linda Bankkauffrau *und* aktives Mitglied in der Frauenbewegung ist (Tversky & Kahneman, 1983). Die nach Tversky und Kahneman (1983) irrelevante Zusatzinformation, dass Linda Philosophie-Graduierte ist und sich für soziale Rechte und Umweltfragen einsetzt, führte aber bei 85% der Probanden von Tversky und Kahneman (1983, S. 299) dazu, dass sie die Konjunktion als wahrscheinlicher angaben.

Problematisch ist die „Vagheit“ (Gigerenzer, 1996) der Heuristiken des „Heuristics and Biases“ Programms (Gilovich et al., 2002; Kahneman, Slovic & Tversky, 1982). Es fehlt ein Modell über die kognitiven Prozesse einer Heuristik, aus dem neue empirisch prüfbar Hypothesen abgeleitet werden können (Gigerenzer, 1996, S. 593). Es ist nicht klar, wie Informationen einer Entscheidungsaufgabe und deren Repräsentation (Gigerenzer & Hoffrage, 1995) genutzt werden, um zu einer Entscheidung zu gelangen.⁶ Es ist weiterhin nicht spezifiziert, unter welchen Bedingungen Heuristiken angewendet bzw. *nicht* angewendet werden (Gigerenzer, 1996).

Die verwendeten Entscheidungsaufgaben sind weiterhin so konstruiert, dass eine Entscheidung nach einer Heuristik zu einer falschen Antwort führt (Lopes, 1991). Dies ist auch logisch notwendig, um zu zeigen, dass Entscheider systematisch, d. h. in Richtung der Heuristik, von der normativ korrekten Antwort abweichen (Lopes, 1991, S. 73). Es können also Aussagen darüber gemacht werden, ob Menschen in die Richtung einer Heuristik entscheiden. Das bedeutet aber nicht zwangsläufig, dass Entscheider die Heuristiken, die Gilovich et al. (2002) vor-

⁵Dieser Name wurde in der Originalstudie verwendet (Tversky & Kahneman, 1983).

⁶siehe auch Lewandowsky (1993) für die Vorteile einer Software-Implementierung einer genau spezifizierten, mathematischen Theorie

1. Theorie

schlagen, auch tatsächlich anwenden. Eine normative Beurteilung einer Heuristik für die verwendete Aufgabe ist zulässig. Die Anwendung der Repräsentativitäts-Heuristik führt ja zu einer falschen Entscheidung beim oben beschriebenen Linda Problem.⁷ Eine generalisierte Aussage, d. h. über Entscheidungsumwelten hinweg, über die Nützlichkeit bzw. Fehleranfälligkeit der Heuristik ist jedoch nicht erlaubt.

Es kann durch den Fokus auf die Fehleranfälligkeit von Heuristiken ein verzerrtes Bild über die Nützlichkeit von Heuristiken entstehen (Gigerenzer, 1996).⁸ Um die Nützlichkeit einer Heuristik zu bestimmen, ist es jedoch notwendig zu erfassen, wie oft eine Heuristik über ein Set von unterschiedlichen Entscheidungsaufgaben zu korrekten Entscheidungen führt. Die Repräsentativität einer Aufgabe für eine Entscheidungsaufgabe aus dem Alltag kann bestimmt werden (d. h. z. B., wie häufig diese Aufgabe auftritt oder wie wahrscheinlich die Kombination der Merkmale der Aufgabe für die Realität ist), um die praktische Relevanz einer Aufgabe abzuschätzen (Brunswik, 1955; Dhimi, Hertwig & Hoffrage, 2004). Wichtig ist hierbei jedoch, dass zur Bewertung der Relevanz einer Aufgabe die Repräsentativität allein kein ausreichendes Kriterium ist (Jekel, Fiedler & Glöckner, 2011): Die Konsequenz eines (irrationalen) Verhaltens in einer unrepräsentativen - hier im Sinne von seltenen - Aufgabe kann drastisch ausfallen (z. B. in der Krisenbewältigung einer Atom- oder Umweltkatastrophe) oder eine seltene Aufgabe (bzw. die Einkleidung einer Aufgabe) kann überproportional oft zu ihrem Auftreten in einer natürlichen Umwelt präsentiert werden, um so irrationales Verhalten auszunutzen (z. B. in der Werbung).

Daraus ergibt sich die Frage, für welche Art von Problemen eine Heuristik nützlich bzw. weniger nützlich ist. Es geht also um die Passung einer Heuristik zu einer (Entscheidungs-)Umwelt (Gigerenzer & Goldstein, 1996; Gigerenzer et al., 1999). Todd und Gigerenzer (2000, S. 741) unterscheiden zwischen einem Kohärenz- und Korrespondenzkriterium der Rationalität (siehe auch Hammond & Stewart, 2001; Newell, 2005); ob eine Heuristik rational ist, lässt sich daran festmachen, ob sie zu einem normativen Standard wie der Wahrscheinlichkeitstheorie kohärent ist (Tversky & Kahneman, 1974) und/oder mit akkuraten, schnellen und informationssparsamen Entscheidungen in einer Umwelt

⁷siehe hierzu aber Gigerenzer (1996, S. 593), der die Angemessenheit der Vergleichsnormen, die sich aus der Wahrscheinlichkeitstheorie ableiten lassen, auf die Entscheidungsaufgaben von Tversky und Kahneman (1974) in Frage stellt

⁸siehe Lopes (1991) für eine Analyse der von Tversky und Kahneman (1974) verwendeten Rhetorik, die nahelegt, dass auch die Autoren diesem Fehlschluss aufsitzen

korrespondiert (Gigerenzer & Goldstein, 1996).⁹ Die Nützlichkeit einer Heuristik für eine Entscheidungsumwelt kann also auch als Rationalitätskriterium verwendet werden. Die Güte einer Heuristik nach dem Korrespondenzkriterium wird auch die *ökologische* Rationalität einer Heuristik genannt (Goldstein & Gigerenzer, 2002).

Eine Strategie, die das Kohärenzkriterium erfüllt, muss nicht das Korrespondenzkriterium erfüllen. So zeigen Martignon und Schmitt (1999, S. 589), dass eine Strategie, die auf der optimalen Suchreihenfolge der Cues basiert, im Vergleich zu weniger computational aufwändigen Strategien (wie *Take The Best*, siehe weiter unten) schlechter abschneidet, wenn die Methode der Kreuzvalidierung angewendet wird, d. h., wenn die optimale Suchreihenfolge in einem Lernset von Entscheidungsaufgaben ermittelt wird und in einem unabhängigen Testset auf ihre Performanz hin getestet wird. Um die Rationalität von Heuristiken zu beurteilen, kann das Korrespondenzkriterium - die Ermittlung der Akkuratheit von Heuristiken in einer Entscheidungsumwelt - Aufschluss darüber geben, welche Heuristiken „gut“ sind, und das Kohärenzkriterium - der Vergleich der Prozesse von Heuristiken mit einem normativen Standard - warum diese gut sind (Newell, 2005, S. 14).

Nach Simon (1956, S. 133) können sich Entscheider über einfache „mechanische Wahlmechanismen“, die den begrenzten kognitiven Eigenschaften eines Organismus Rechnung tragen, adaptiv an die strukturellen Umwelteigenschaften anpassen, um „annähernd“ rational zu entscheiden. Um begrenzte Rationalität zu erforschen, ist es also notwendig, die „Prozesse“ eines Systems und die Struktur der Umwelt, für die sie adaptiv sind, zu bestimmen (Simon, 1990, S. 6). Menschliches rationales Verhalten ist demnach geformt durch eine „Schere“, deren Schneideklingen die „rechnerischen Fähigkeiten eines Handelnden“ und die „Umweltstruktur einer Aufgabe“ sind (Simon, 1990, S. 7).

Menschliche Entscheider optimieren ihre Entscheidungen nicht nach klassisch rationalen Standards, sondern beenden die Suche nach Alternativen, wenn eine Entscheidungsalternative - z. B. ein bestimmter Lebenspartner (Gigerenzer & Goldstein, 1996, S. 651) - die „Erwartungen basierend auf Erfahrungen“ darüber, wie „angemessen“ gut eine Entscheidungsalternative sein kann, erfüllt (Simon, 1990, S. 9). Ein

⁹Kohärenz kann in Bezug auf eine Heuristik nur graduell sein: Eine Heuristik kann per Definition nicht kohärent zu einer normativen Lösung sein, da eine Heuristik weniger Information für ein Urteil heranzieht und/oder die Verrechnung der Cues vereinfacht (Gigerenzer & Gaissmaier, 2011) und somit immer nur eine Annäherung an eine normative Lösung sein kann.

1. Theorie

begrenzt rationaler Entscheider wird also nicht nach allen möglichen Entscheidungsalternativen - allen möglichen Partnern in der Gegenwart und Zukunft - suchen, den Wert - wie z. B. mit einem bestimmten Partner glücklich zu werden - und die Wahrscheinlichkeit des Eintreffens der Konsequenz, die mit der Entscheidungsalternative verbunden sind, bestimmen und den Erwartungswert für jede Entscheidungsalternative berechnen (Bernoulli, 1738), sondern die erste Entscheidungsalternative wählen, die seinem Anspruch genügt. Ein Entscheidungsalgorithmus, der dies leistet, wird „satisficing“ Heuristik genannt (siehe Simon, 1990, S. 9 und Gigerenzer & Goldstein, 1996, S. 651).

Der Zusammenhang zwischen einem Entscheider und seiner Umwelt lässt sich im Brunswikschen Linsenmodell darstellen (Brunswik, 1955, 1956; Hammond, 1966; Hammond & Stewart, 2001); siehe Abbildung 1.1 (S. 23). Die Ausprägung des distalen Kriteriums ist latent, sie entzieht sich der direkten Beobachtung. Die Ausprägung der distalen Variablen manifestiert sich aber in den Cues, die die Umwelt hergibt.

In welchem Maße die Cues mit der distalen Variablen zusammenhängen, spiegelt sich in der ökologischen Validität der Cues wieder. Die Cues werden wiederum von einem Entscheider für seine zentrale Antwort (central response) resp. Urteil (Gigerenzer & Kurz, 2001, Fig. 24.1, S. 345) über den Zustand der distalen Variablen genutzt. In welchem Maße Cues genutzt werden, spiegelt sich in der Cue-Nutzung, das ist der Zusammenhang zwischen den Cues und der Entscheidung, wieder.

Sollen z. B. die Einwohnerzahlen von Städten geschätzt werden und sind diese nicht direkt beobachtbar (latente distale Variable), so kann ein Entscheider Cues, die mit der Einwohnerzahl von Städten zusammenhängen (ökologische Validität), wie z. B., ob eine Stadt eine Universität hat oder nicht, in einer bestimmten Weise nutzen (Cue-Nutzung), um auf die Einwohnerzahlen der Städte zu schließen (zentrale Antwort).

Statistisch lässt sich die ökologische Validität und die Nutzung von Cues über Korrelationen erfassen oder, um dem Konzept des „vicarious functioning“ (Brunswik, 1955, S. 207) - d. h. der Redundanz von Cues durch Interkorrelation - gerecht zu werden, über eine Regression mit den Cues als Prädiktoren und den Ausprägungen von distalen Variablen bzw. den Entscheidungen über die Ausprägungen der distalen Variablen als Kriterium berechnen.

Ein akkurater Entscheider nutzt die validen Cues für eine Entscheidung über die Ausprägung einer distalen Variablen. Statistisch lässt sich das über die Passung zwischen der Cue-Nutzung und den ökologi-

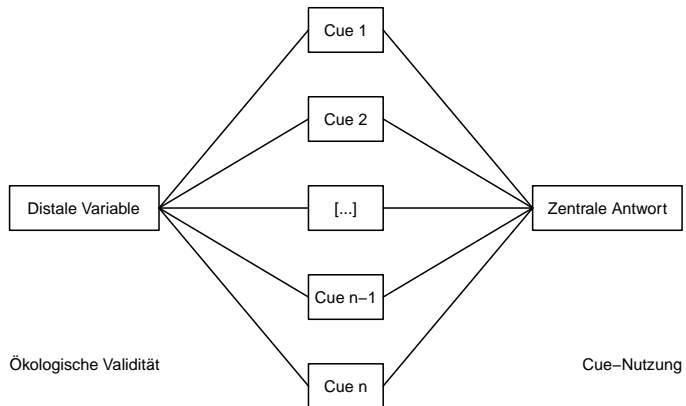


Abbildung 1.1. Das Brunswiksche Linsenmodell (vgl. Brunswik, 1955, Fig. 8, S. 206). Der Entscheider schließt anhand von Cues auf den Zustand der distalen Variablen. Ökologische Validität und Cue-Nutzung beschreiben den Zusammenhang zwischen der distalen Variablen, den Cues und der zentralen Antwort (resp. Urteil).

schen Validitäten berechnen.¹⁰

In der Tradition von Simon (1956) und Brunswik (1955) beschäftigt sich das Forschungsprogramm der adaptiven Werkzeugkiste (Gigerenzer et al., 1999) mit „schnellen und (informations-)sparsamen“ Heuristiken einer Entscheidung, das sind die Werkzeuge der Werkzeugkiste eines Entscheiders, und deren Adaptivität für Entscheidungsumwelten.

Eine Entscheidungsheuristik aus dem Forschungsprogramm der adaptiven Werkzeugkiste sollte dabei folgende Charakteristiken aufweisen (Gigerenzer & Goldstein, 1996; Gigerenzer & Todd, 1999): Die drei Prozess-Bausteine einer Entscheidung, das sind die Suche nach Cues,

¹⁰siehe z. B. Gosling, Ko, Mannarelli und Morris (2002) für ein illustratives Anwendungsbeispiel des Brunswikschen Linsenmodells auf die Wahrnehmung von Persönlichkeit

1. Theorie

der Abbruch der Suche nach Cues und die Integration der Cues zu einer Entscheidung (Gigerenzer & Goldstein, 1996, S. 653), sollen für Heuristiken spezifiziert sein. Entscheidungsheuristiken sollen mathematisch genau beschrieben sein (Gigerenzer & Todd, 1999, S. 17) und mit Blick auf die kognitiven Fähigkeiten eines Entscheiders einfach, d. h. wenig rechenintensiv, sein.

Entscheidungsheuristiken sollen ein mittleres Maß an „Spezifität“ aufweisen (Gigerenzer & Todd, 1999, S. 19), d. h., sie dürfen nicht zu unspezifisch für eine Entscheidungsumwelt sein, da sie sonst relevante Eigenschaften einer Umwelt nicht erfassen, aber auch nicht zu spezifisch auf eine Entscheidungsumwelt zugeschnitten sein, da sie sonst irrelevanten Eigenschaften einer Umwelt erfassen.¹¹ Nur wenn ein mittlere Maß an Spezifität gelingt, ist eine Generalisierung der Entscheidungsheuristik, d. h. eine Anwendung der Heuristik auf neue Daten aus der Umwelt, erfolgreich und somit robust gegen die Fehlervarianz der Stichprobe an Daten aus der Umwelt (Gigerenzer & Todd, 1999, S. 20), an der die Parameter der Heuristik angepasst wurden; siehe dazu auch die Anmerkungen zur Methode der Kreuzvalidierung auf S. 33 in Kapitel 1.2.4.

In Bezug auf das Brunswiksche Linsenmodell (Brunswik, 1955) kann auf der rechten Seite des Linsenmodells, auf der die Entscheidungen modelliert werden, auch eine Heuristik stehen. Wie akkurat eine Heuristik ist, ergibt sich daraus, ob die Heuristik die relevanten Cues, das sind die Cues mit einer hohen ökologischen Validität, korrekt zu einer Entscheidung integriert (Nutzung der Cues).

Das Vorgehen im Forschungsprogramm der adaptiven Werkzeugkiste besteht zusammenfassend aus vier Schritten (Gigerenzer & Todd, 1999, S. 16): 1.) Computermodelle von Entscheidungsheuristiken, die die oben beschriebenen Charakteristiken haben, werden implementiert, 2.) für jede Entscheidungsheuristik wird die Umweltstruktur, für die die jeweilige Entscheidungsheuristik adaptiv ist, analysiert, 3.) für jede Entscheidungsheuristik wird die Akkuratheit der Entscheidungsheuristik in Entscheidungsumwelten ermittelt und 4.) für jede Entscheidungsheuristik wird erfasst, ob Menschen die Entscheidungsheuristik auch tatsächlich anwenden.

¹¹siehe hierzu auch das „bias-variance dilemma“ (Gigerenzer & Brighton, 2009, S. 117) und das nicht ganz ernst gemeinte „polynomial law of sensation“ von Nihm (1976), das so viele freie Parameter wie Datenpunkte - 1 hat und somit immer 100% der Varianz in den Daten „aufklären“ kann

1.2. Entscheidungsheuristiken und Umwelten

Analog zur Forschungsstrategie der adaptiven Werkzeugkiste (siehe letzter Abschnitt im vorherigen Kapitel) wird in diesem Kapitel auf die Parameter (Kapitel 1.2.1, S. 25) und Architektur (Kapitel 1.2.2, S. 27; Kapitel 1.2.3, S. 30) von Entscheidungsheuristiken und auf deren Akkuratheit in Abhängigkeit zu Entscheidungsumwelten (Kapitel 1.2.4, S. 31) eingegangen. Es wird exemplarisch anhand der prominenten *Take The Best* Heuristik (Kapitel 1.2.5.1, S. 42) und der Rekognitions-Heuristik (Kapitel 1.2.5.2, S. 46) die Evidenz zu der Hypothese, dass Menschen diese Entscheidungsheuristiken auch anwenden (Kapitel 1.2.5, S. 38), dargestellt.

1.2.1. Parameter eines Cues

Ob Stadt A oder Stadt B größer ist, kann in der Städtepaarvergleichsaufgabe von Gigerenzer und Goldstein (1996) anhand von (z. B.) folgenden Cues geschlossen werden: Ob Stadt A oder B Hauptstadt ist, Messestadt ist, eine Fußballmannschaft in der ersten Liga hat, eine IC Zugsanbindung hat, Landeshauptstadt ist, ein einstelliges Autokennzeichen besitzt, eine Universität hat, einen Industriekomplex hat oder in Ostdeutschland liegt. Die Eigenschaften eines Cues können anhand von drei Merkmalen, der Validität, der Diskriminanzrate und der Successrate, beschrieben werden (Martignon & Hoffrage, 1999); siehe Tabelle 1.1 (S. 26).

Die Validität eines Cues gibt an, in wie vielen Entscheidungen der Cue proportional zu allen Entscheidungen, in denen der Cue zwischen den Optionen diskriminiert, auf die korrekte Option zeigt. Wenn z. B. der Hauptstadt-Cue zwischen den Städten diskriminiert, das ist dann der Fall, wenn ein Vergleich zwischen Berlin und einer anderen Stadt gezogen wird, deutet er auf die größere Stadt. Berlin ist die einzige Hauptstadt im Set und die Stadt mit der höchsten Anzahl von Einwohnern. Der Hauptstadt Cue hat demnach eine Validität von 1.

Eine Validität von .5 bedeutet für einen Paarvergleich, dass der Cue in der Hälfte der Fälle auf die korrekte und in anderen Hälfte auf die falsche Option zeigt. Eine Entscheidung basierend auf einem Cue von .5 kann somit niemals besser sein als eine Entscheidung basierend auf Raten, d. h., der Cue hat keinen Informationswert.

Wenn ein Cue eine Validität unter .5 hat, deutet das Nicht-Vor-

1. Theorie

Tabelle 1.1. Validitäten, Diskriminanzraten und Successraten für alle neun Cues der Städtepaarvergleichsaufgabe (vgl. mit Gigerenzer & Goldstein, 1996, Table 1, S. 655).

| Cue Name | Validität | Diskrimi- nanzrate | Successrate |
|---------------------------------|-----------|-----------------------|-------------|
| Hauptstadt | 1.00 | .02 | .51 |
| Messestadt | .91 | .28 | .62 |
| Fußballmannschaft 1. Liga | .87 | .30 | .61 |
| IC Zuanbindung | .78 | .38 | .61 |
| Landeshauptstadt | .77 | .30 | .58 |
| Einstelliges Autokennzeichen | .75 | .34 | .59 |
| Universität | .71 | .51 | .61 |
| Industriekomplex | .56 | .30 | .52 |
| Ost-Deutschland | .51 | .27 | .50 |

Anmerkungen: Die Parameter basieren auf dem von Anna Kuhlen and Lael Schooler revalidierten Städte-Cues-Set.

handensein des Cue-Merkmals auf die korrekte Option hin. Die Validität eines Cues von z. B. .1 wird dann zu .9 umkodiert: In 90% der Fälle deutet das Nicht-Vorhandensein des Cue-Merkmals auf die korrekte Option. Die Validität eines Cues kann also im Paarvergleich Werte zwischen .5 und 1 annehmen.

Die Diskriminationsrate gibt an, in wie vielen Entscheidungen proportional zu allen Entscheidungen ein Cue zwischen den Optionen diskriminiert, d. h., die Ausprägungen des Cues für die Vergleiche unterschiedlich ist. Der Hauptstadt Cue hat z. B. zwar eine hohe Validität, er kann aber nur angewendet werden, wenn ein Vergleich zwischen Berlin und einer anderen Stadt gezogen wird. Bei einem Set von 83 Städten - und damit 3,403 möglichen Vergleichen - sind das 82 Städtevergleiche, was einer Diskriminanzrate von $82/3403 \approx .02$ entspricht. Die Diskriminanzrate eines Cues kann im Paarvergleich Werte zwischen 0 und $\approx .5$ annehmen.

Es ist kritisch anzumerken, dass die Validität, wie sie bei Gigerenzer und Goldstein (1996) definiert wird, blind für die Anzahl der Verglei-

che in einer Entscheidungsaufgabe ist, die in die Berechnung einfließen. So hat ein Cue eine Validität von 1 unabhängig davon, ob er nur in einer Entscheidung auf die korrekte Option deutet und in 149 weiteren Vergleichen nicht diskriminiert oder ob er in allen 150 Vergleichen auf die korrekte Option deutet (Lee & Cummins, 2004, S. 345). In der Successrate eines Cues werden die Validität *und* die Diskriminanz eines Cues verrechnet. Die Formel lautet (Newell, Rakow, Weston & Shanks, 2004, Formel 1, S. 120): $(\text{Diskriminanzrate} \times \text{Validität}) + ((1 - \text{Diskriminanzrate}) \times .5)$. Die Successrate eines Cues kann im Paarvergleich Werte zwischen .5 und $\approx .75$ annehmen.

Im Folgenden werden beispielhaft Heuristiken beschrieben, in denen - basierend auf den Cue-Parametern - in einer bestimmten Weise Cues konsultiert und verrechnet werden. Man kann zwischen Entscheidungen basierend auf einem Cue (Kapitel 1.2.2, S. 27) und Entscheidungen basierend auf mehr als einem Cue (Kapitel 1.2.3, S. 30) unterscheiden.

1.2.2. ORDM: Entscheidungen basierend auf einem Cue

Ein Anwender der *Take The Best* Heuristik (Gigerenzer & Goldstein, 1996, 1999) *sucht* die Cues nach der Höhe ihrer Validität ab, *stoppt* die Cue Suche nach dem ersten Cue, der zwischen den Optionen diskriminiert, und *entscheidet* in Richtung dieses Cues. Wenn kein Cue zwischen den Optionen diskriminiert, wählt ein Anwender von *Take The Best* eine der Optionen zufällig. Wenn einem Entscheider also z. B. die Cues aus Tabelle 1.2 (S. 28) für den Vergleich zwischen Stuttgart und Nürnberg vorliegen, wird er zuerst den validesten Cue, das ist der Hauptstadt Cue, dann den zweit validesten Cue, das ist der „Fußballmannschaft 1. Liga“ Cue, konsultieren. Da der Hauptstadt Cue und „Fußballmannschaft 1. Liga“ Cue¹² nicht zwischen den Optionen Stutt-

¹²An diesem Cue lässt sich verdeutlichen, dass sich die Validität eines Cues mit der Zeit auch verändern kann. Welche Stadt am Ende einer Fußball-Saison noch eine Mannschaft in der 1. Liga hat, wird zumindest im unteren Tabellenbereich der 1. Liga variieren. Ein weiteres Beispiel für die Veränderung der Validität von Cues über die Zeit ist der Rekognitions-Cue im Kontext der Vorhersage von Aktiengewinnen: Aktiengewinne aus Portfolios, die aus Firmen bestanden, die in einer Umfrage mindestens 90% der Teilnehmer bekannt waren, fielen in einer Studie von Borges, Goldstein, Ortman und Gigerenzer (1999) höher aus als die Gewinne von Portfolios basierend auf Expertenmeinungen und Marktindizes. Boyd (2001) konnte jedoch empirisch zeigen, dass die Eigenschaften des Marktes, die zum Zeitpunkt der Studie von Borges et al. (1999) bestanden und zeitlich hoch instabil sind, ursächlich für den positiven Zusammenhang zwischen

1. Theorie

gart und Nürnberg diskriminiert, führt erst der Landeshauptstadt Cue zu einer korrekten Entscheidung, da Stuttgart im Gegensatz zu Nürnberg Landeshauptstadt ist und tatsächlich mehr Einwohner hat. Ein *Take The Best* Entscheider wählt also in diesem Vergleichsfall korrekterweise Stuttgart als die Stadt, die mehr Einwohner hat.

Tabelle 1.2. Beispiel für ein Cue-Muster für vier Cues für den Vergleich zwischen Stuttgart und Nürnberg (vgl. mit Gigerenzer & Goldstein, 1996, Anhang, S. 668).

| Cue Name | Städtevergleich | |
|---------------------------|-----------------|----------|
| | Stuttgart | Nürnberg |
| Hauptstadt | 0 | 0 |
| Fußballmannschaft 1. Liga | 1 | 1 |
| Landeshauptstadt | 1 | 0 |
| Universität | 1 | 1 |

Anmerkungen: 1 bedeutet Merkmal liegt vor, 0 bedeutet Merkmal liegt nicht vor.

Die *Take The Best* Heuristik beinhaltet den Rekognitions-Cue (vgl. mit Gigerenzer & Goldstein, 1996, Fig. 2, S. 653), dem im Forschungsprogramm der adaptiven Werkzeugkiste eine besondere Rolle zukommt (Goldstein & Gigerenzer, 2002). Der Rekognitions-Cue wird vor allen anderen Cues konsultiert. Wenn eine der Optionen bekannt ist und die andere Option nicht, dann wählt ein Entscheider, der den Rekognitions-Cue verwendet, die Option, die bekannt ist. Wenn der Rekognitions-Cue nicht diskriminiert, dann bedeutet das entweder, dass beide Städte bekannt sind und möglicherweise Wissen über die Städte, wie z. B. die oben angeführten Cues, abgerufen werden kann, oder dass keine der beiden Städte bekannt ist und somit auch kein weiteres Wissen über die Städte vorliegt und geraten werden muss. Damit der Rekognitions-Cue valide sein kann, muss ein Zusammenhang zwischen der Rekognition eines Objekts und dem Entscheidungskriterium vorliegen. Für die Städtevergleichsaufgabe ist das der Fall (Goldstein & Gigerenzer, 2002): Die Größe einer Stadt korreliert mit der Rekognitionsrate einer

Rekognitionsraten und Aktiengewinnen und damit für die hohen Gewinne durch die Anwendung des Rekognitions-Cue waren.

Stadt und wird über die Häufigkeit, mit der über Städte in den Medien berichtet wird, in der Umwelt vermittelt. Über Städte mit vielen Einwohnern wird tendenziell mehr in den Medien berichtet, was wiederum die Wahrscheinlichkeit erhöht, dass diese Städte bekannt sind. Dies ließ sich auch empirisch zeigen (Goldstein & Gigerenzer, 2002): Für eine amerikanische Probanden-Stichprobe lag die Korrelation zwischen der Einwohnerzahl der untersuchten deutschen Städte und der Rekognitionsrate der Städte bei .6, die Korrelation zwischen der Rekognition und der Häufigkeit der Nennung der Städte in der Chicago Tribune bei .79 und der Korrelation zwischen der Nennung in der Chicago Tribune und der Einwohnerzahl der Städte bei .7. Der Rekognitions-Cue hat eine weitere wichtige Eigenschaft: Er kann nur angewendet werden, wenn Objekte in dem Set unbekannt sind. Weniger Wissen über Städte kann also zu besseren Entscheidungen führen, wenn der Rekognitions-Cue valider als die Wissens-Cues (siehe Tabelle 1.1, S. 26) ist und genügend häufig zwischen den Optionen diskriminiert.¹³

Ein Anwender der *Take The Last* Heuristik (Gigerenzer & Goldstein, 1996, 1999; Todd & Gigerenzer, 2000) konsultiert für die erste Entscheidungsaufgabe einen Cue in zufälliger Reihenfolge. Ab der zweiten Entscheidungsaufgabe wird der Cue konsultiert, der in der vorangegangenen Entscheidung diskriminiert hat. Wenn dieser Cue nicht diskriminiert, wird die Suchreihenfolge der Cues per Zufall bestimmt und die Cuesuche erst abgebrochen, wenn ein Cue diskriminiert oder alle Cues konsultiert wurden und eine Option per Zufall gewählt werden muss (Gigerenzer & Goldstein, 1999, S. 80). Die *Take The Last* Heuristik maximiert also die Wahrscheinlichkeit, dass nur ein Cue für eine Entscheidung konsultiert werden muss (Newell et al., 2004, S. 120). Im Gegensatz zu der *Take The Best* Heuristik muss ein Entscheider bis auf die Richtung der Cues, d. h., welche Cue Ausprägung mit einem höheren Entscheidungskriterium einhergeht, nicht die genauen Validitäten der Cues kennen, sondern nur die Information, welcher Cue zuletzt diskriminiert hat, im Gedächtnis behalten. Die *Take The Last* Heuristik ist also im Gegensatz zu der *Take The Best* Heuristik weniger rechenintensiv.

Ein Anwender der *Minimalist* Heuristik (Gigerenzer & Goldstein, 1996, 1999) wählt die Suchreihenfolge der Cues zufällig und entschei-

¹³Die Anwendung des Rekognitions-Cues wird in späteren Studien (Goldstein & Gigerenzer, 2002; Marewski, Gaissmaier, Schooler, Goldstein & Gigerenzer, 2010) die „Rekognitions-Heuristik“ genannt. Hier zeigt sich eine gewisse Beliebigkeit in der Erfindung „neuer“ Heuristiken.

1. Theorie

det in Richtung des ersten Cues, der diskriminiert. Im Gegensatz zu der *Take The Best* und der *Take The Last* Heuristik muss ein Entscheider nur die korrekte Richtung der Cues und nicht die genaue Validität der Cues oder den zuletzt diskriminierenden Cue kennen. Die *Minimalist* Heuristik ist demnach im Gegensatz zu der *Take The Best* Heuristik und der *Take The Last* Heuristik weniger rechenintensiv. *Minimalist* kann als Vergleichsheuristik zu *Take The Best* genutzt werden, um zu überprüfen, wie stark die Information über die Validität von Cues die Akkuratheit einer Heuristik erhöht (bzw. möglicherweise sogar vermindert).

1.2.3. MRDM: Entscheidungen basierend auf mehreren Cues

Ein Anwender der *Tallying* Heuristik (Gigerenzer & Goldstein, 1996) sucht alle verfügbaren Cues ab und integriert diese in eine Entscheidung. Dabei werden für jede Option die Cues, die für die Option sprechen, aufaddiert, die Summen zwischen den Optionen verglichen und die Option mit der größeren Summe gewählt. Wenn die Summe gleich hoch ist, wird eine der Optionen per Zufall gewählt. Im Beispiel (Tabelle 1.2, S. 28) sprechen drei Cues für Stuttgart und zwei Cues für Nürnberg, so dass ein Entscheider die korrekte Option Stuttgart wählt. *Dawes Rule*¹⁴ (Gigerenzer & Goldstein, 1999) ist eine leichte Abwandlung von *Tallying*: Die Cues, die nicht für die Option sprechen - in Tabelle 1.2 (S. 28) sind das alle 0 Positionen - werden von der Summe der Cues, die für die Option sprechen, abgezogen.¹⁵ *Dawes Rule* führt im Vergleich zu *Tallying* zu identischen Ergebnissen, so lange alle Cue Ausprägungen bekannt sind.

Ein Anwender der *Weighted Tallying* Heuristik (Gigerenzer & Goldstein, 1996) wählt die Option, die eine höhere Summe über die mit den Validitäten gewichteten Cues aufweist. Wenn die Summe gleich hoch ist, wird eine der Optionen per Zufall gewählt. Im Beispiel - vgl. Tabelle 1.2 (S. 28) und Tabelle 1.1 (S. 26) - gibt das einen Wert von $0 \times 1 + 1 \times .87 + 1 \times .77 + 1 \times .71 = 2.35$ für Stuttgart und einen Wert von 1.58 für Nürnberg und führt zu der korrekten Entscheidung, dass

¹⁴Die Bezeichnung *Dawes Rule* geht auf die Forschung von Robyn Dawes (1979) zurück, der die Robustheit von Unit-Weight Modellen im Vergleich zur multiplen Regression, in der die Evidenz mit den Betas des Modells gewichtet wird, untersucht hat.

¹⁵Bei Gigerenzer und Goldstein (1996) wird *Dawes Rule* noch unter dem Namen *Unit-Weight Linear Model* geführt.

Stuttgart die Stadt mit mehr Einwohnern ist. *Franklin's Rule*¹⁶ (Gigerenzer & Goldstein, 1999) ist eine leichte Abwandlung von *Weighted Tallying*: Die Summe über die gewichteten Cues, die nicht für die Option sprechen, werden von der Summe über die gewichteten Cues, die für die Option sprechen, abgezogen. *Franklin's Rule* führt im Vergleich zu *Weighted Tallying* zu identischen Ergebnissen, wenn alle Cue Ausprägungen bekannt sind.¹⁷

1.2.4. Überblick Studien zur normativen Güte

In den Studie von Gigerenzer und Goldstein (1996, 1999) wurden die oben angeführten ORDM und MRDM Entscheidungsstrategien auf ihre Akkuratheit in der schon beschriebenen Städtevergleichsaufgabe untersucht. In der Städtevergleichsaufgabe sind basierend auf dem Rekognitions-Cue und neun Wissens-Cues (siehe Tabelle 1.1, S. 26) 3,403 mögliche Paarvergleiche zwischen 83 Städten zu machen und zu entscheiden, welche Stadt im jeweiligen Paarvergleich mehr Einwohner hat.

Ergänzend wurde die multiple lineare Regression, die zu den MRDM Entscheidungsstrategie gezählt werden kann, für den Strategievergleich mit aufgenommen. In der multiplen Regression wird von den Differenzmustern der Cues der Paarvergleiche als Prädiktoren auf die Differenzen der Städtegrößen im Kriterium geschlossen. Im Vergleich zu den beschriebenen MRDM Heuristiken beinhalten die Gewichte einer multiplen linearen Regression die Information über die Interkorrelation der Cues. Die multiple lineare Regression zieht daher in Relation zu den anderen Entscheidungsstrategien die meiste Information aus der Umwelt und kann als interessanter Vergleich für die Akkuratheit der weniger rechenintensiven Heuristiken verwendet werden (Gigerenzer & Goldstein, 1996, S. 658).

Die Rekognitionsrate, d. h., wie viele Städte bekannt sind, und das Wissen über die Ausprägungen der Cues wurden systematisch variiert. Es wurden jeweils 500 Probanden pro Entscheidungsstrategie mit jeweils einer Rekognition von 0 bis 83 Städten und 0%, 10%, 20%, 50%, 75% und 100% Wissen über die Cues der Städte, über die Rekognition besteht, simuliert. Dabei bedeutet z. B. 10% Cue Wissen, dass in der

¹⁶Bei Gigerenzer und Goldstein (1996) wird *Franklin's Rule* noch unter dem Namen *Weighted Linear Model* geführt. Es ist unverständlich, warum es zu diesen Inkonsistenzen in der Namensgebung bei den selben Autoren kommt.

¹⁷Da in den Studien dieser Arbeit nur vollständige Cue-Patterns verwendet werden, wird hier auf die Forschung zu Inferenzmechanismen bei fehlender Information nicht eingegangen (siehe hierzu z. B. Garcia-Retamero & Rieskamp, 2008, 2009).

1. Theorie

Matrix von 83 Städten \times 9 Wissens Cues = 747 Positionen per Zufall 90% \approx 672 Positionen für die Strategien nicht zugänglich sind. Für die jeweilige Rekognitionsrate wurden die Städte, über die Rekognition besteht, so gewählt, dass der Rekognitions-Cue eine Validität von .8 hatte. Das bedeutet, wenn der Rekognitions-Cue anwendbar war, hat er in 80% der Vergleiche auch zu einer korrekten Entscheidung geführt. Der Wert von .8 basiert auf einer empirischen Untersuchung der Validität der Rekognition von realen Probanden über die Städte der Städtevergleichsaufgabe (Gigerenzer & Goldstein, 1996, S. 656).

Die Parameter der untersuchten Entscheidungsstrategien¹⁸, d. h. Validitäten, Diskriminanzraten und Regressionsgewichte der Cues, wurden über die Vergleiche aller 83 Städte ermittelt. Insgesamt wurden also für jede Entscheidungsstrategie 500 simulierte Probanden \times 84 Rekognitionsraten \times 6 Raten an Cue Wissen = 252,000 Akkuratheitsraten berechnet.

Die Entscheidungsstrategien zeigen ähnlich hohe Akkuratheitsraten: Im Mittel über die jeweils 500 simulierten Probanden für alle Raten an Cue Wissen und Rekognition liegt die höchste Akkuratheitsrate bei .658 für *Take The Best*, d. h., *Take The Best* trifft im Mittel über alle Entscheidungen hinweg 65.8% korrekte Entscheidungen, und die niedrigste Akkuratheitsrate bei .645 für *Take The Last* (Gigerenzer & Goldstein, 1996, Table 2, S. 660). *Take The Best* schneidet ungefähr genauso gut wie die rechenintensive *Multiple Regression* ab, bei der die mittlere Akkuratheitsrate bei .657 liegt. Gleichzeitig konsultieren simulierte *Take The Best* Nutzer im Mittel über alle Bedingungen aber nur drei Cues (Gigerenzer & Goldstein, 1999, Table 4-2, S. 87) im Gegensatz zu allen möglichen 10 Cues (1 Rekognitions-Cue + 9 Wissens-Cues = 10 Cues), die die kompensatorischen Strategien für eine Entscheidung konsultieren *und* integrieren.

Wenn alle Cue-Ausprägungen bekannt sind, d. h., 100% Cue Wissen vorliegt, steigen die Akkuratheitsraten für *Take The Best*, *Weighted Tallying* und *Tallying* und *Multiple Regression* mit steigender Anzahl an Städten, über die Rekognition besteht. Die höchste Akkuratheitsrate von .78 wird jedoch bei einer Rekognition zwischen 60 und 65 Städten erreicht (Gigerenzer & Goldstein, 1996, Figure 6, S. 659). Weniger Wissen über die Städte führt also zu höheren Akkuratheitsraten, da nur dann der valide Rekognitions-Cue genutzt werden kann, wenn

¹⁸Zur Erinnerung: das sind *Take The Best*, *Take The Last*, *Minimalist*, *Tallying*, *Dawes Rule*, *Weighted Tallying*, *Franklin's Rule* und *Multiple Regression*; siehe Kapitel 1.2.2 (S. 27) und Kapitel 1.2.3 (S. 30).

mindestens eine und weniger als alle Städte bekannt sind; Gigerenzer und Goldstein (1996, S. 656) nennen das einen „less-is-more“ Effekt. *Dawes Rule* und *Franklin's Rule* zeigen bei mittlerer Rekognitionsrate sogar einen Einbruch in den Akkuratheitsraten von circa 3% (Gigerenzer & Goldstein, 1996, Figure 6, S. 659), was darauf zurückzuführen ist (Gigerenzer & Goldstein, 1996, S. 660), dass bei einer Stadt, über die Rekognition besteht, eine kompensatorische Entscheidungsstrategie auch die neun Wissens-Cues integriert und in der Städteaufgabe das Merkmal eines Cues oft nicht vorliegt und damit in der Anwendung von *Dawes Rule* und *Franklin's Rule* als negative Evidenz für die bekannte Stadt, die aufgrund der hohen Validität des Rekognitionscues meist die Stadt mit mehr Einwohner ist, gezählt wird.

Simulierte *Take The Best* Entscheider kommen also schneller zu einer Entscheidung und treffen im Vergleich zu rechenintensiven Algorithmen wie *Multiple Regression* dennoch gleich viele korrekte Entscheidungen. Ein Unterschied zwischen den Akkuratheitsraten von *Take The Best* und *Multiple Regression* zeigt sich (Gigerenzer & Goldstein, 1999), wenn die Methode der Kreuzvalidierung (Browne, 2000) angewandt wird. Bei der Kreuzvalidierung werden die Parameter einer Entscheidungsstrategie anhand einer zufälligen Stichprobe aller Städte - dem sogenannten Lernset - ermittelt bzw. „gefittet“. Die Akkuratheit der Strategien wird dann am unabhängigen Set der übrigen Städte - dem sogenannten Testset - ermittelt. Das Fitten der Parameter und die Messung der Akkuratheit wird also an unabhängigen Städtesets durchgeführt. Eine Entscheidungsstrategie, die Fehlervarianz modelliert, wird, sofern der Fehler des Lernsets und des Testsets unabhängig voneinander und unsystematisch sind, keine besseren Akkuratheitsraten als der Zufall erzielen, d. h. Raten zwischen den Optionen. Über die Methode der Kreuzvalidierung wird also die Generalisierbarkeit der Entscheidungsstrategien auf ein unabhängiges Städteset untersucht (Gigerenzer & Goldstein, 1999). Die Größe des Lernsets und damit die Größe des Testsets kann als experimenteller Faktor variiert werden.

Bei einer Lernsetgröße von 10% der Städte erreicht *Multiple Regression* eine Akkuratheitsrate im Testset von $\approx .61$ und *Take The Best* von $\approx .66$ (Gigerenzer & Goldstein, 1999, Figure 4-3a, S. 93). Je größer das Lernset wird, desto geringer wird der Unterschied der Akkuratheitsraten zwischen *Take The Best* und *Multiple Linear Regression*. Ab einer Lernsetgröße von 60% aller Städte überdecken sich die Akkuratheitskurven beider Entscheidungsstrategien. Die Akkuratheitsraten gemittelt über ein Set von 20 unterschiedlichen Inhaltsdomänen von

1. Theorie

probabilistischen Entscheidungsaufgaben¹⁹ ergeben zwei Befunde (Gigerenzer et al., 1999; Martignon & Hoffrage, 1999). *Take The Best* erzielt eine hohe Akkuratheitsrate von $\approx .72$ im Testset (Martignon & Hoffrage, 1999) und ist somit also nicht nur eine gute Heuristik für die Städtevergleichsaufgabe. Die Güte einer Heuristik (nach dem Korrespondenzkriterium) ergibt sich dabei aus dem Vergleich zu anderen Entscheidungsstrategien. So zeigte z. B. die *Multiple Regression* ein typisches Overfitting Muster in den Akkuratheitsraten: *Multiple Regression* erzielt im Lernset eine höhere Akkuratheitsrate von $\approx .77$ als *Take The Best* von $\approx .75$, die Akkuratheitsrate fällt jedoch im Testset für *Multiple Regression* auf $\approx .68$ ab (Martignon & Hoffrage, 1999, Figure 6-5, S. 135).

Take The Best erzielt selbst im Vergleich zu Algorithmen aus der Forschung zur künstlichen Intelligenz - wie *CART*, *C4.5*, *Nearest Neighbor Classifier* und einem *Perceptron* Netzwerkmodell - höhere Akkuratheitsraten im Testset in der Städtevergleichsaufgabe (Gigerenzer & Brighton, 2009, Figure 2, S. 115).

Es stellt sich die Frage, warum eine einfache Heuristik wie *Take The Best* vergleichbar gut oder besser als rechenintensivere Strategien abschneidet (Gigerenzer & Brighton, 2009; Hogarth & Karelaia, 2006, 2007; Martignon & Hoffrage, 1999, 2002). Die Gründe liegen in der Architektur der Heuristik und der Umwelt (Hogarth & Karelaia, 2007), in der sie angewendet wird. MRDM Strategien können nicht akkurater als ORDM Heuristiken wie *Take The Best* sein, wenn das Entscheidungskriterium aus einer nonkompensatorischen Verrechnung der Cues in einer Umwelt besteht (Martignon & Hoffrage, 2002, Theorem 3, S. 47). Wenn es sich um binäre Cues - 1 oder 0 für Merkmal liegt vor oder Merkmal liegt nicht vor - handelt, bedeutet nonkompensatorisch (Hogarth & Karelaia, 2006, Formel 3, S. 212), dass das Gewicht eines Cues - gemeint sind die Gewichte einer Linearkombination über die Cues als Prädiktoren des Entscheidungskriteriums - größer ist als die Summe der geringeren Gewichte von Cues. In diesem Fall ergibt die Integration von Cues in einer MRDM Strategie keinen Informationsvorteil, da die geringer prädiktiven Cues den prädiktivsten Cue

¹⁹Die Aufgaben stammen aus so unterschiedlichen Bereichen wie der Psychologie - die Einschätzung der Attraktivität von berühmten Männern basierend auf der Beliebtheit, der Bekanntheit und der Nationalität als Cues - der Ökonomik - die Einschätzung von Professorengelältern basierend auf dem Geschlecht, der Position, dem höchsten Abschluss und den Berufsjahren als Cues - und Biologie - die Einschätzung der Anzahl der gelegten Eier von arktischem Saibling basierend auf dem Gewicht, dem Alter und dem mittleren Eigewicht als Cues.

nicht überstimmen können. *Take The Best* zeigt gute Akkuratheitsraten, selbst wenn die Validität der Cues fehlerbehaftet ist, so lange der valideste Cue identifiziert werden kann (Hogarth & Karelaia, 2006); *Take The Best* benötigt also keine exakten Validitäten, was das Erlernen einer erfolgreichen Suchreihenfolge für Cues weniger aufwendig macht (Todd & Dieckmann, 2005). *Take The Best* erzielt jedoch auch in kompensatorischen Umwelten zu MRDM Strategien vergleichbare Akkuratheitsraten, d. h., wenn die Cues der Umwelt redundant sind, d. h., hoch miteinander korreliert sind, und somit die Verrechnung von mehr Cues nicht mehr Information bedeutet (Hogarth & Karelaia, 2007).²⁰ In der Kreuzvalidierung erzielt *Take The Best* im Vergleich zu MRDM Strategien im Testset höhere Akkuratheitsraten, weil *Take The Best* sensitiv genug ist, um wahre Varianz im Kriterium der Entscheidungsumwelt zu modellieren, ohne zu sensitiv zu sein und Fehlervarianz zu modellieren (Gigerenzer & Brighton, 2009).

Problematisch ist die Auswahl der Objekte für die Simulation einer Entscheidungsumwelt. Um die Validität eines Cues zu interpretieren, muss die Referenzklasse, über die die Validität berechnet wurde, angegeben werden (Hoffrage & Hertwig, 2006). In der Städtevergleichsaufgabe (Gigerenzer & Goldstein, 1996, 1999) befinden sich in der Referenzklasse alle deutschen Städte, die mehr als 100,000 Einwohner haben. Der Schwellenwert von 100,000 Einwohnern ist jedoch arbiträr. So zeigt sich, dass ein Referenzklasse, die nur aus deutschen Städten besteht, die mehr als 300,000 Einwohner haben, zu anderen Cue Validitäten führt (Hoffrage & Hertwig, 2006). Das kann zu starken Effekten führen: Bei einer Reduktion der Größe der Referenzklasse durch eine Schwellenänderung von 100,000 zu 300,000 Einwohnern sinkt z. B. die Validität des „Fußballmannschaft 1. Liga“-Cues von .9 auf .65 ab; es kann sogar passieren, dass sich die Cue Ausprägung, die auf eine größere Stadt hindeutet, umdreht, wie das z. B. bei dem „Industriekomplex“-Cue der Fall ist (Hoffrage & Hertwig, 2006, Table 16.1, S. 393). Es stellt sich dann die Frage, welche Referenzklasse die „adäquate“ (Hoffrage & Hertwig, 2006, S. 401) Referenzklasse für eine Simulation ist, wenn es das Ziel ist, die Referenzklasse realer Entscheider im Labor abzubilden.

Hoffrage und Hertwig (2006) greifen das Konzept der Repräsentativität (Brunswik, 1955, 1956; Dhimi et al., 2004) auf. Repräsentativität bedeutet dabei, dass Objekte in Laborstudien verwendet werden sollen, mit denen Entscheider in einer realen Entscheidungsumwelt kon-

²⁰siehe zur Redundanz von Cues auch die Anmerkungen zu „vicarious functioning“ (Brunswik, 1955) auf S. 22

1. Theorie

frontiert sind und für die ihre - somit inhaltspezifischen (Cosmides & Tooby, 1994) - Heuristiken geeicht und damit adaptiv sind (Gigerenzer et al., 1999). Für das Labor sind danach also die Aufgaben geeignet, die auch in der realen Welt (häufig) vorkommen. Dazu soll aus der Population aller möglichen Objekte eine *repräsentative Stichprobe* (Hoffrage & Hertwig, 2006, S. 402) gezogen werden. So können z. B. die Entscheidungsumwelten von Probanden untersucht werden, um Objekte, mit denen Entscheider konfrontiert werden, zu identifizieren. Hoffrage und Hertwig (2006) geben aber nicht an, wie das z. B. für die Städtevergleichsaufgabe praktisch funktionieren soll. Eine weitere Möglichkeit für die Gewinnung einer repräsentativen Stichprobe ist die Nutzung von Mediatoren in der Umwelt - wie z. B. der Häufigkeit, mit der über eine Stadt in der Presse berichtet wird - als Proxy für die Wahrscheinlichkeit zu verwenden, dass ein Objekt in die Referenzklasse realer Entscheider fällt (Hoffrage & Hertwig, 2006, S. 402). Es ist jedoch dann unverständlich, dass in den bisherigen Studien (Czerlinski et al., 1999; Gigerenzer & Brighton, 2009; Gigerenzer & Goldstein, 1996, 1999; Katsikopoulos, Schooler & Hertwig, 2010; Martignon & Hoffrage, 1999) die Wahrscheinlichkeit, dass ein Objekt in einer Simulation in das Lernset fällt, für alle Städte gleich hoch gesetzt wurde (Dougherty, Franco-Watkins & Thomas, 2008; Jekel, in Vorbereitung): Der hohe Methodenanspruch wird also selbst in den eigenen Studien nicht erfüllt.²¹

An den vorgestellten Studien lässt sich weiterhin kritisieren, dass das Zufallsniveau von .5 nicht von den Validitäten der Cues abgezogen wurde (Bröder, 2000c; Glöckner & Betsch, 2008b; Jekel & Glöckner, in Vorbereitung). Ein Cue mit einer Validität von .5 hat keinen Informationsgehalt: Er führt wie eine durchschnittliche Zufallsentscheidung in einer Paarvergleichsaufgabe in 50% der Fälle zu einer korrekten und in 50% der Fälle zu einer falschen Entscheidung. Durch die Korrektur erhält ein Cue mit einer Validität von .5 ein Gewicht von 0, was dazu führt, dass die Ausprägung des Cues in der Verrechnung des Cues in einer MRDM Strategie *ignoriert* wird.²² Durch die Korrektur wird weiterhin sichergestellt, dass ein valider Cue nicht durch wenig valide Cues in der unkorrigierten gewichteten Cue-Summe überstimmt werden kann.

²¹siehe Hertwig, Hoffrage und Martignon (1999) als einzige Ausnahme zu Sampling-Wahrscheinlichkeiten in Relation zur Stadtgröße in der Simulation zur *QuickEst* Heuristik

²²korrigierte Validität \times Cueausprägung $\rightarrow 0 \times$ Cueausprägung = 0

Unkorrigierte Cue Validitäten können bei MRDM Strategien zu irrationalen Entscheidungen führen. Angenommen es liegt z. B. ein Cue-Muster von 1 0 0 für Option 1 und 0 1 1 für Option 2 vor und die Validitäten der Cues sind 1, .51 und .51, dann ergibt die gewichtete Cue-Summe (d. h. also *Weighted Tallying*) der unkorrigierten Cues einen Wert von 1 für Objekt 1 und 1.02 für Objekt 2 und die gewichtete Cue-Summe der korrigierten Cues einen Wert von .5 für Objekt 1 und .02 für Objekt 2. Nur die korrigierten Validitäten stellen also in diesem Fall eine rationale Entscheidung für Option 2 sicher: Ein Cue mit einer Validität von 1 ist per Definition in allen Entscheidungen, in denen er diskriminiert, korrekt. Die Integration von zufallskorrigierten Validitäten führt aber nicht immer zu der normativ korrekten Lösung, dies ist im Falle der Verrechnung probabilistischer Cues über das Bayesianische Schlussfolgern erreichbar (siehe z. B. Lee & Cummins, 2004). So kann z. B. ein Cue mit einer Validität von .98 durch 49 Cues mit einer Validität von .51 bei *Weighted Tallying* überstimmt werden; nach Bayes sollte dies erst ab einer Anzahl von 98 Cues der Fall sein - gegeben die Entscheidungsoptionen sind a-priori gleich wahrscheinlich die korrekte Lösung und die Cues sind unabhängig voneinander. MRDM basierend auf zufallskorrigierten Validitäten ist also im Kontext einer probabilistischen Entscheidung eine Heuristik (Jekel & Glöckner, in Vorbereitung).

Studien zur normativen Güte, die mit unkorrigierten Cue Validitäten rechnen (Czerlinski et al., 1999; Gigerenzer & Brighton, 2009; Gigerenzer & Goldstein, 1996, 1999; Martignon & Hoffrage, 1999), benachteiligen MRDM Strategien, was besonders dann kritisch ist, wenn sie als Vergleich für ORDM Strategien, die nicht durch diesen methodischen *Fehler* beeinflusst werden, herangezogen werden. Das ist z. B. der Fall für die *Take The Best* Heuristik, bei der die Subtraktion der Ratewahrscheinlichkeit von den Cuevaliditäten keinen Einfluss auf die Suchreihenfolge hat. Selbst für den Fall, dass ein Cue von .5 konsultiert wird, weil alle valideren Cues nicht diskriminieren, kommt die Konsultation des Cues einem durchschnittlichen Rateprozess gleich, der auch per default bei *Take The Best* implementiert ist, wenn keiner der Cues diskriminiert.

Auch für Studien zur deskriptiven Güte, d. h. Studien in denen die Anwendungshäufigkeit der Strategien von Probanden in einer Entscheidungsaufgabe untersucht wird, sind unkorrigierte Validitäten problematisch. Die Anzahl der Nutzer einer MRDM Strategie wird unterschätzt, wenn (1) die verwendeten Entscheidungsaufgaben unterschiedliche Vor-

1. Theorie

hersagen für eine korrigierte und unkorrigierte MRDM Strategie treffen, (2) die korrigierte MRDM Strategie psychologisch plausibler ist (also eher von Probanden angewendet wird) und (3) sich die korrigierte MRDM Strategie nicht im Set der Strategien befindet, für die die Passung zwischen Vorhersage und beobachtetem Entscheidungsverhalten bestimmt wird (siehe z. B. Garcia-Retamero & Rieskamp, 2009; Mata, Schooler & Rieskamp, 2007; Mata, Helversen & Rieskamp, 2010, 2011; Rieskamp, 2006; Rieskamp & Hoffrage, 2008; Rieskamp & Otto, 2006). Die Situation wird weiterhin verschärft, wenn Wahlen, die konsistent zu einer korrigierten MRDM Strategie sind, konsistent zu einer Alternativstrategie sind und somit fälschlicherweise als Evidenz für die Anwendung der Alternativstrategie gewertet werden. So klassifizierten z. B. Rieskamp (2006) und Rieskamp und Hoffrage (2008) Wahlen von Probanden als Wahlen, die konsistent zu *Take The Best* oder *Franklin's Rule*²³ mit unkorrigierten Validitäten waren. In einer Simulationsstudie (Jekel & Glöckner, in Vorbereitung) ließ sich jedoch für die in der Studie von Rieskamp (2006) und Rieskamp und Hoffrage (2008) verwendeten Parameter (Anzahl und Validitäten der Cues und Anzahl der verwendeten Aufgaben) zeigen, dass Wahlen, die von simulierten Probanden unter Anwendung von *Franklin's Rule* mit korrigierten Validitäten produziert worden waren, bis zu 48% als *Take The Best* Wahlen (mis-)klassifiziert wurden. Das heißt also, die ausschließliche Testung von MRDM Strategien mit unkorrigierten Validitäten kann zur Überschätzung der Nutzung von ORDM Strategien führen (Jekel & Glöckner, in Vorbereitung).

1.2.5. Überblick Studien zur deskriptiven Güte

Eine typische probabilistische Entscheidungsaufgabe, die bei der Untersuchungen zur Anwendung von Entscheidungsstrategien verwendet wird, ist das Börsenspiel (Bröder, 2000b, 2000c, 2003; Bröder & Schiffer, 2003b, 2006a, 2006b; Newell & Shanks, 2003; Newell et al., 2004; Rakow, Newell, Fayers & Hersby, 2005)²⁴; siehe Abbildung 1.2 (S. 39)

²³Rieskamp (2006) und Rieskamp und Hoffrage (2008) nennen die Strategie in Anlehnung an Payne et al. (1993) *Weighted Additive*, so dass es in der Literatur (unbefriedigenderweise) drei unterschiedliche Namen für die gleiche Strategie gibt: *Weighted Linear Model* (Gigerenzer & Goldstein, 1996), *Franklin's Rule* (Gigerenzer & Goldstein, 1999) und *Weighted Additive*.

²⁴siehe z. B. Bröder und Newell (2008) und Bröder (2011) für einen informativen Überblick zu empirischen Arbeiten im Forschungsprogramm der Werkzeugliste und der *Take The Best* Heuristik im Speziellen

1.2. Entscheidungsheuristiken und Umwelten

für einen Screenshot eines Durchgangs des Börsenspiels, wie ihn die Probanden am Computer sehen. Das Paradigma zur Erfassung von Entscheidungsverhalten wurde von Payne et al. (1988) unter dem Namen *Mouselab*-Studie entwickelt.

The screenshot shows a web-based interface for a stock game. At the top left, it says 'Aufgabe Nr. 1'. The main area is titled 'Angebote:' and contains a table with four rows of cues and three columns for 'Aktie 1', 'Aktie 2', and 'Aktie 3'. The cues are: 'Umsatzsteigerung?', 'Gewinnsteigerung?', 'Dividendenerhöhung?', and 'Neue Standorte eröffnet?'. The first cell in the first row (Aktie 1, Umsatzsteigerung) contains the text 'ja', while all other cells contain a question mark. Below the table, there is a yellow box labeled 'Wähle:' containing three buttons: 'Aktie 1', 'Aktie 2', and 'Aktie 3'. To the right of this box is another yellow box labeled 'Bereits investiert:' containing the text '-2 % des Gewinns'.

| | Aktie 1 | Aktie 2 | Aktie 3 |
|---------------------------|---------|---------|---------|
| Umsatzsteigerung ? | ja | ? | ? |
| Gewinnsteigerung ? | ? | ? | ? |
| Dividendenerhöhung ? | ? | ? | ? |
| Neue Standorte eröffnet ? | ? | ? | ? |

Wähle:

Aktie 1 Aktie 2 Aktie 3

Bereits investiert: -2 % des Gewinns

Abbildung 1.2. Die Oberfläche eines 3 Optionen x 4 Cues Börsenspiels. Ein Proband setzt 2% des möglichen Gewinns für die Konsultation des Cues „Umsatzsteigerung“ für Aktie 1 ein.

In der Aufgabe kann ein Proband eine Aktien von drei Aktien auswählen. Die Aufgabe ist es, die Aktie mit dem höchsten Gewinn zu wählen und somit den Kontostand über ein Set von Durchgängen zu maximieren.

Es gibt Cues, im Beispiel sind das vier Cues - ob die betreffende Firma eine Umsatzsteigerung, eine Gewinnsteigerung und eine Dividendenerhöhung hatte und ob neue Standorte eröffnet wurden -, die mit einer bestimmten Validität mit dem Kriterium, das ist der Anstieg des Gewinns der Aktie, zusammenhängen. Die Cue Ausprägungen sind verdeckt. Eine Cue Ausprägung, im Beispiel sind das binäre Cues („ja“ versus „nein“), kann durch Anklicken des entsprechenden Optionen-Cue Feldes eingeholt werden. Die Konsultation eines Cues ist mit Kosten verbunden, im Beispiel sind das pro Cue 2% Abzug von dem Gewinn der Aktie.

Die Validität und Richtung der Cues wird den Probanden mitgeteilt

1. Theorie

oder die Probanden müssen diese erlernen (Newell, Weston & Shanks, 2003; Rieskamp, 2008). Nach jedem Durchgang gibt es ein Feedback oder kein Feedback (Bröder, 2000b); eine beispielhafte Feedback Oberfläche für den drei Optionen Fall des Börsenspiels ist in Abbildung 1.3 (S. 40) dargestellt.

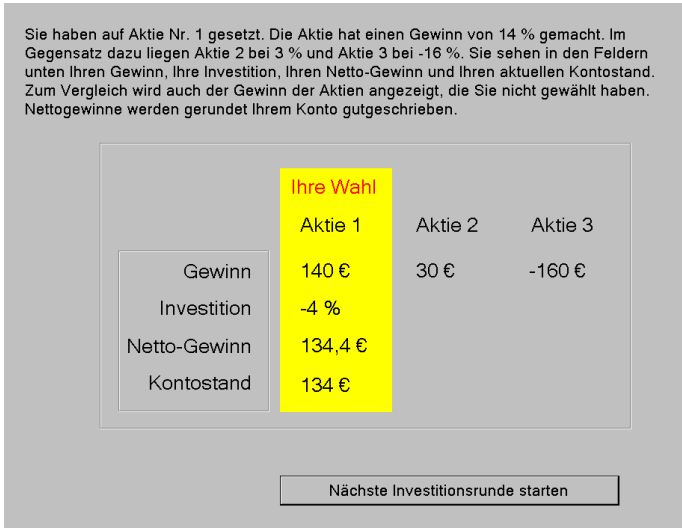


Abbildung 1.3. Die Oberfläche der Feedback Maske eines drei Optionen Börsenspiels. Angezeigt wird der Gewinn jeder Aktie, die Investition, die ein Proband in einem Durchgang durch die Konsultation von Cues gemacht hat, wie viel Gewinn die Aktie nach Abzug der Investition gemacht hat und der aktuelle Kontostand.

Um festzustellen, welche Strategie ein Entscheider anwendet, lässt sich das beobachtete Entscheidungsverhalten in einer Aufgabe wie dem Börsenspiel mit dem nach einer Strategie erwarteten Entscheidungsverhalten vergleichen. Es ist dann möglich, die Passung der Modelle zu dem beobachteten Entscheidungsverhalten über statistische Indizes wie einer Fehlerabweichung (z. B. root mean squared deviation), dem Bayesianischen Informations-Kriterium (Schwarz, 1978) oder dem Akaike Informationskriterium (Akaike, 1974; Burnham & Anderson, 2002) zu bestimmen.²⁵ Entscheidungsverhalten kann die Entscheidung (Bröder

²⁵siehe aber zum Beispiel Myung (2000) und Pitt, Myung und Zhang (2002) zu den

& Schiffer, 2003a) selbst, alle Variablen, die mit einer Entscheidung einhergehen, wie zum Beispiel die Konfidenz in eine Entscheidung (siehe z. B. Glöckner & Betsch, 2008a), und die Prozesse (Payne et al., 1988, 1993), die zu einer Entscheidung führen, sein, wie z. B. die benötigte Entscheidungszeit, die Anzahl der konsultierten Cues oder die mit einem Eye-Tracker aufgezeichneten Suchmuster (Glöckner & Herbold, 2011; Horstmann, Ahlgrimm & Glöckner, 2009) in einer probabilistischen Entscheidungsaufgabe.²⁶

Bröder (2000a) kritisiert, dass die Art der Integration von Cues nicht zwangsläufig mit den Entscheidungs-Prozessen korrespondieren muss. So kann ein Proband, der strikt in Richtung des ersten diskriminierenden validesten Cues entscheidet - ein erwartetes Entscheidungs-Muster für *Take The Best* -, trotzdem vor der Entscheidung alle vorhandenen Cues konsultieren - ein erwartetes Suchmuster für MRDM Strategien. Es ist unklar, wie mit einer Diskrepanz in den Messvariablen umgegangen werden soll, wenn es a) keine Fehlertheorie darüber gibt, wie unsystematische Abweichungen zwischen erwartetem und beobachtetem Entscheidungsverhalten zu einem Fehler-Index verrechnet werden können (Bröder, 2000a, S. 654) und es b) keinen Ansatz dazu gibt, wie die Fehler über alle Maße hinweg gewichtet und zu einem Gesamt-Index verrechnet werden können, um so für jede Strategie einen einzigen Passungs-Index für den Vergleich zwischen Strategien berechnen zu können.

Als Weiterentwicklung der Bayesianischen Strategieklassifikation (Bröder, 2010; Bröder & Schiffer, 2003a), über die die Passung einer Strategie zu den Daten anhand der Entscheidungen bestimmt wird, ermöglicht die *Multiple-Measure Maximum Likelihood* Strategieklassifikations-Methode (Glöckner, 2009, 2010; Jekel et al., 2010), verschiedene Entscheidungsmaße in *einem* Passungs-Index für eine Strategie zu verrechnen, wenn sich aus den Strategien genaue mathematische Vorhersagen zu den Entscheidungsmaßen für die verwendeten Aufgaben ableiten lassen, (hoch-)diagnostische Entscheidungsaufgaben verwendet werden (Jekel et al., 2011) und die zugrundeliegenden Verteilungen der Entscheidungsmaße bekannt sind (Jekel et al., 2010, S. 56). Bei Entscheidungen, die strategiekonform oder -inkonform sein können, ist das die Binomialverteilung; an die Verteilung der logarithmierten Entscheidungszeiten und der Konfidenzurteile kann eine Normalverteilung

Problemen mit diesen Indizes

²⁶siehe Glöckner und Witteman (2010) für einen Überblick zu den Messmethoden für Entscheidungsprozesse

1. Theorie

approximiert werden (Glöckner, 2009; Jekel et al., 2010); siehe Anhang A.1 (S. 167) für die mathematischen Grundlagen der Bayesianischen Strategieklassifikation.

Es können über die Messung der Prozesse von Entscheidungen und der Methode der Strategieklassifikation von Entscheidern in einer Entscheidungsumgebung wie dem Börsenspiel experimentelle Einflussfaktoren auf die Strategiewahl untersucht werden.

1.2.5.1. *Take The Best* Heuristik

Probanden einer *Mouselab* Studie zu Lotterie-Präferenzaufgaben (Payne et al., 1988), in der der Zeitdruck und die Varianz der Cue-Validitäten intraindividuell über die Durchgänge der Aufgabe variiert wurden, zeigten auf den Prozessmaßen einer Entscheidung Suchverhalten, das mit ORDM einhergeht: Wenn der Zeitdruck erhöht wurde und die Varianz der Cue Validitäten groß war, trafen die Probanden schneller eine Entscheidung, konsultierten weniger Cues, verbrachten mehr Zeit mit der Konsultation des validesten Cues und deckten die Felder im *Mouselab* mehr cue- als optionsweise auf, d. h., dass die Optionen-Cues-Matrix, wie sie z. B. in Abbildung 1.2 (S. 39) angeordnet ist, tendenziell eher zeilen- als spaltenweise abgesucht wird.

In einer Studie von Bröder (2000b, Experiment 4, S. 1342) führten hohe Suchkosten dazu, dass die Anzahl der *Take The Best* Nutzer von $\approx 23\%$ auf 65% anstieg (Bröder, 2000b, Figure 1, S. 1343). In einer Studie von Newell und Shanks (2003, S. 57), in der die Cue Validitäten von vier Cues in einer Trainingsphase für Paarvergleiche im Börsenspiel gelernt werden mussten, konnte der Befund zu den Suchkosten repliziert werden: 75% versus 13% der Probanden zeigten Entscheidungsverhalten konsistent zu *Take The Best*, wenn die Suchkosten hoch versus niedrig waren. Weiterhin führten Hinweise auf die Rang-Ordnung der Cue-Validitäten und der Grad, in dem die Entscheidungsumwelt deterministisch²⁷ waren, zu Entscheidungsverhalten konsistent zu *Take The Best*. Newell und Shanks (2003, S. 63) konnten jedoch beobachten, dass 50% der Probanden in 50% der relevanten Entscheidungen mehr als einen diskriminierenden Cue konsultierten und damit die Stopp-Regel der *Take The Best* Heuristik verletzten.

Die Befunde zur Komplexität einer Entscheidungs-Aufgabe fallen wi-

²⁷Der Grad des Determinismus einer Umwelt kann über die Höhe des zufälligen Fehlerterms ϵ in der (Linear-)kombination von Cues manipuliert werden, d. h. Kriterium = $\beta_0 + \beta_1 \times \text{Cue}_1 + \dots + \beta_n \times \text{Cue}_n + \epsilon$.

dersprüchlich aus. Bei einem Anstieg der Anzahl der Optionen und Cues tendieren Probanden dazu, prozentual weniger Cues aufzudecken (Payne, 1976, Table 4, S. 381) und zu einer cueweisen Suche über Alternativen zu wechseln (Table 5, S. 383). Der Autor schließt, dass mehr Informationen zu nonkompensatorischem Entscheidungsverhalten führen. An der Studie ist jedoch zu kritisieren, dass der Autor eine relativ zu der Anzahl der dargebotenen Informationen berechnete Anzahl an aufgedeckten Informationen als abhängige Variable verwendet, die die absolute Anzahl der tatsächlich aufgedeckten Informationen verschleiert. So berichtet Payne (1976, S. 381), dass die absolute Anzahl der von den Probanden inspierten Cues zwar anstieg, jedoch in Relation zu der Anzahl der gegebenen Informationen mit steigender Anzahl an Informationen im *Mouselab* sank. Newell et al. (2003) konnten dagegen zeigen, dass die Erhöhung der Anzahl von vier Cues, die in den Studien zum Börsenspiel zuvor standardmäßig verwendet wurden (S. 84), auf 6 Cues nur bei 33% der Probanden (S. 86) zu einem Entscheidungsverhalten konsistent zu *Take The Best* führte, wenn man die korrekte Befolgung der Such-, Stopp- und Entscheidungsregel als Klassifikationskriterium verwendet, und „nur“ 46% der Probanden (S. 88) überhaupt eine Heuristik verfolgten, in der die Suche nach dem ersten diskriminierenden Cue abgebrochen wurde. Möglicherweise ist die deterministische Befolgung der Entscheidungsregeln ein zu strenges Klassifikationskriterium, so dass die Verwendung der Bayesianischen Strategieklassifikation (Bröder & Schiffer, 2003a; Glöckner, 2009) als Methode hier angebrachter wäre.

Es zeigt sich über 12 Datensätze zum Börsenspiel hinweg (Bröder & Newell, 2008, Figure 1, S. 208), dass Entscheider sensibel für die Struktur der Umwelt sind. Das Verhältnis der in einer Umwelt zu erwartenden Gewinne für die Nutzung von *Take The Best* zu *Franklin's Rule* korreliert positiv über die Studien hinweg mit einem Wert von Pearson's $r = .83$ (S. 208) mit dem Prozentsatz der Probanden, die als *Take The Best* Anwender klassifiziert wurden. Die hohe Varianz in der Nutzung der *Take The Best* über die Studien hinweg zeigt weiterhin, dass die Anwendung der *Take The Best* Heuristik als „Universal-Hypothese“ (Bröder, 2000b, S. 1335), d. h. *Take The Best* wird von allen Entscheidern immer verwendet, verworfen werden muss. Es ist jedoch - a-priori - fraglich, wie eine Universalhypothese überhaupt in das Forschungsprogramm der adaptiven Werkzeugkiste passt: Bei Gültigkeit der Universalhypothese (Bröder, 2000b, 2011) wäre die Anwendung der *Take The Best* Heuristik unabhängig von der Umwelt und damit

1. Theorie

wenig adaptiv.

Es zeigt sich jedoch auch, dass nicht alle Probanden eine Heuristik wählen, die adaptiv zur Umwelt ist. In den Bedingungen, in denen die Gewinne für die Nutzung der *Take The Best* Heuristik höher ausfallen als für die Nutzung von *Franklin's Rule*, verwendeten $\approx 49\%$ der Probanden *nicht* die *Take The Best* Heuristik (Bröder & Newell, 2008, Figure 1, S. 208).²⁸

Es zeigt sich weiterhin, dass Probanden Schwierigkeiten haben, adaptiv auf eine sich verändernde Umweltstruktur zu reagieren: Bröder und Schiffer (2006a) konnten zeigen, dass Probanden im Börsenspiel dazu tendierten, eine einmal gewählte adaptive Strategie beizubehalten, selbst wenn sich über die Aufgaben hinweg, in der Studie war das nach 80 Börsenspiel-Durchgängen, die Umwelt verändert und die Anwendung der Strategie zu geringeren Gewinnen führt und somit nicht mehr adaptiv ist.

Es zeigte sich weiterhin in einer Studie von Bröder (2003), dass die Verwendung der *Take The Best* Heuristik nicht durch die Belastung des Arbeitsgedächtnisses, was über die Notwendigkeit Informationen der Optionen-Cues-Matrix memorieren zu müssen manipuliert wurde, ansteigt. Wenn jedoch wie in einer Studie von Bröder und Schiffer (2003b) die Cue-Muster für Optionen *vor* der Vergleichsphase zwischen Optionen gelernt werden und somit in der Vergleichsphase komplett aus dem Gedächtnis abgerufen werden müssen, zeigt sich, wie in der Studie bei 64% der Probanden (S. 283) beobachtet, eine erhöhte Nutzung der *Take The Best* Heuristik.

Darüber hinaus konnte Bröder (2003) zeigen, dass das Repräsentationsformat der Cues einen Einfluss auf die Strategiewahl bei Entscheidungen aus dem Gedächtnis haben: So sank die Nutzung der *Take The Best* Heuristik für Entscheidungen aus dem Gedächtnis mit steigendem Grad an bildhafter Präsentation der Optionen und Cues von 47.4% auf 21.1% (Table 1, S. 284). Je nach Cue-Format lässt sich sogar zeigen, dass Entscheider komplett von einer Anwendung regelbasierter (non-)kompensatorischer Entscheidungs-Strategien abweichen. In einer Studie von Bröder, Newell und Platzter (2010) mussten die Probanden in einer fiktiven medizinischen Paarvergleichs-Aufgabe bestimmen, welcher Patient einen kritischeren Krankheitszustand hat. Dazu mussten die Probanden in einer vorgeschalteten Phase die Symptommuster für

²⁸Der Wert ergibt sich aus der Differenz von 100% - dem Mittelwert über die Ordinaten-Werte der Kreis-Datenpunkte, die einen Wert größer 1 auf der Abszisse in Figure 1 (Bröder & Newell, 2008, S. 208) haben.

jeden Patient memorieren und in der Phase der Paarvergleiche abrufen. Es zeigt sich bei einem binären Cue Format, in dem ein Cue einen von zwei inhaltsorthogonalen Symptomen, wie z. B. Fieber versus Unterkühlung, annehmen konnte, im Vergleich zu einem „klassischen“ Cue Format, bei dem ein Symptom vorliegen oder nicht vorliegen kann, eine erhöhte Anwendung von exemplarbasierten Strategien (Juslin & Persson, 2002). In der Studie führte die Darstellung der Cues als alternative Symptome zu 39% versus 6% Nutzern (S. 335) einer exemplarbasierten Strategie.²⁹

Es zeigt sich über mehrere Studien hinweg (Bröder, 2003, 2005), dass die Nutzung der *Take The Best* Heuristik nicht mit bestimmten Persönlichkeitsmerkmalen eines Entscheider einhergeht. Es konnten keine Zusammenhänge zwischen der Wahl einer Heuristik und den Persönlichkeits-Dimensionen der Big Five (Borkenau & Ostendorf, 1993) gefunden werden (Bröder, 2005, Tabelle 8-4, S. 158). In einer nonkompensatorischen Umweltstruktur des Börsenspiels zeigte sich (Bröder, 2003), dass Probanden mit einer höheren verbalen Intelligenz (Jäger, Süß & Beauducel, 1997) eher die adaptive *Take The Best* Heuristik verwendeten. Spannenderweise konnte jedoch kein Zusammenhang mit numerischer Intelligenz (Jäger et al., 1997) gefunden werden, so dass es möglich ist, dass die verbal intelligenteren Probanden eher die Instruktionen der Börsenspielaufgabe verstanden haben und daher eher adaptiv auf die Umwelt reagieren konnten. Gegen diese Interpretation spricht jedoch der unerwartete Befund, dass es *keinen* Zusammenhang zwischen verbaler Intelligenz und der Strategiewahl in einer kompensatorischen Umwelt gab. Über die verbale Intelligenz hinaus hatte die Arbeitsgedächtnis-Spanne eines Probanden keinen Einfluss auf die Strategiewahl in einer Umwelt (Bröder, 2003).

Bröder und Newell (2008, S. 209) schließen, dass die Studien zur *Take The Best* Heuristik Zweifel aufkommen lassen, dass Entscheider, zumindest wenn die Cue Information einfach zugänglich ist, „einfache“ nonkompensatorische Entscheidungs-Heuristiken als erste Wahl („default mode“) verwenden. Die häufig von Entscheidern gezeigte Verletzung der Such-, Stopp- und Entscheidungsregel der *Take The Best* Heuristik und der hohe Anteil an Probanden, der in den Studien nicht als Nutzer der *Take The Best* identifiziert werden konnten, lässt nur den Schluss zu (Newell, 2005, S. 12), dass „manche Leute manchmal“ Entscheidungen konsistent zur *Take The Best* Heuristik treffen. Einige situationale Bedingungen für „manchmal“ wurden in diesem Kapitel genannt. Die

²⁹siehe S. 54 in Kapitel 1.3 für eine kurze Beschreibung von Exemplarmodellen

1. Theorie

persönlichkeitsrelevanten Bedingungen für „manche Leute“ sind jedoch unklar bzw. die in diesem Kapitel genannten Studien weisen nur für Intelligenz einen Zusammenhang mit der adaptiven Strategiewahl auf.³⁰

1.2.5.2. Rekognitions-Heuristik

Der Rekognitions-Cue der *Take The Best* Heuristik hat eine Sonderstellung im Forschungsprogramm der adaptiven Werkzeugkiste, was sich darin zeigt, dass die nonkompensatorische Anwendung des Rekognitions-Cues sogar als *Rekognitions-Heuristik* bezeichnet wird (Goldstein & Gigerenzer, 1999, 2002).

Nach Gigerenzer und Goldstein (1996) wird der Rekognitions-Cue nonkompensatorisch verwendet: Wenn der Cue zwischen zwei Optionen diskriminiert - nur eine von zwei Städten ist bekannt - wird *nur* nach dem Rekognitions-Cue eine Entscheidung *in Richtung* des rekognizierten Objekts getroffen und somit Wissens-Cues der rekognizierten Stadt, sofern über die Rekognition hinaus auch Wissen besteht, ignoriert.

Die Annahme über die nonkompensatorische Verwendung des Rekognitions-Cues führt jedoch allein schon aus theoretischen Überlegungen zu irrationalen Entscheidungen (Newell & Shanks, 2007). In der Städteaufgabe bedeutet das z. B., dass ein Entscheider eine Stadt, von der er *weiß*, dass sie wenige Einwohner hat, gegenüber jeder anderen Stadt, über die er keine Rekognition hat, als größere Stadt wählen muss, wenn Rekognition nonkompensatorisch verwendet wird (Oppenheimer, 2003, S. B3). Dass Entscheider sich in so einem Fall nicht nach der Rekognitions-Heuristik verhalten, zeigt eine Studie von Oppenheimer (2003): Probanden entschieden nur in 37% der relevanten Städtevergleichsaufgaben in Richtung des Rekognitions-Cues, wenn es sich für die Probanden um eine wissentlich kleine Stadt versus eine fiktionale Stadt, die die Probanden also nicht kennen konnten, handelte (Oppenheimer, 2003, Experiment 1, S. B4).

Die Annahme über die fixierte Richtung des Rekognitions-Cues, d. h., ein Entscheider wählt *immer* die Option, über die Rekognition besteht, ist problematisch (Stanovich & West, 2003). Bei Entscheidungen im Konsumentenbereich würde das z. B. bedeuten, dass Entscheider *immer* Markenprodukte kaufen, die teurer aber nicht unbedingt qualitativ bes-

³⁰siehe aber z. B. Hilbig (2008) für den Zusammenhang zwischen Persönlichkeit und der Nutzung der Rekognitions-Heuristik

ser als No-Name Produkte sind³¹, und höhere Kontogebühren bezahlen, da bekannte Banken tendenziell höhere Gebühren verlangen (Stanovich & West, 2003, S. 208). Eine fixierte Richtung des Rekognitions-Cues ist daher nach dem Kriterium der Adaptivität einer Heuristik irrational (Stanovich & West, 2003).³² Es stellt sich die Frage, wie die Nutzung eines Cues adaptiv für eine Umwelt sein kann, wenn sie von der Umwelt losgelöst ist und damit nicht sensibel für unterschiedliche und sich über die Zeit verändernde Umwelten ist.³³

Problematisch ist außerdem die in den Studien zur Rekognitions-Heuristik (Gigerenzer & Brighton, 2009; Goldstein & Gigerenzer, 1999, 2002) verwendete Methodik. So sind Entscheidungen in einer Städtepaarvergleichsaufgabe, die im Schnitt in 90% der relevanten Vergleiche in Richtung der rekognizierten Stadt fallen (Goldstein & Gigerenzer, 2002, S. 81), nicht unbedingt ein Hinweis dafür, dass Entscheider die Rekognitions-Heuristik anwenden, wenn Wissen vorhanden ist, das auf die gleiche Entscheidungsoption deutet. Das ist dann der Fall, wenn der Rekognitions-Cue *und* die Wissens-Cues eine hohe Validität haben und eine mittlere bis hohe Diskriminanzrate und somit per Definition viele Urteilsüberschneidungen aufweisen müssen (Hilbig, 2010a, S. 273). Wenn jedoch die Vorhersage einer Entscheidung nach Rekognition mit der Entscheidung nach Wissen (oder Mischformen zwischen Rekognition und Wissen) in einem Aufgabenset zusammenfällt, sind die Aufgaben nicht geeignet zwischen den Modellen zu diskriminieren. Die Aufgaben sind somit in Bezug auf die Frage, welches Modell Entscheidungsverhalten besser vorhersagen kann, wenig informativ (Glöckner & Betsch, 2011; Jekel et al., 2011). Die Konfundierung zwischen Rekognition und Wissen ist besonders dann schwer zu kontrollieren, wenn

³¹Siehe hierzu auch die Studie von Oeusoonthornwattana und Shanks (2010), die auf eine kompensatorische Verwendung des Rekognitions-Cues in Konsumentenentscheidungen hindeutet.

³²Siehe hierzu auch die Diskussion zur Fluency-Heuristik (Hertwig, Herzog, Schooler & Reimer, 2008; Schooler & Hertwig, 2005), die hier nicht inhaltlich eingeführt werden soll, bei der aber auch eine Entscheidung basierend auf einem Cue in eine fixierte Richtung angenommen wird (Schooler & Hertwig, 2005, S. 612). Unkelbach (2007) konnte jedoch zeigen, dass die Richtung des Fluency Cues flexibel - d. h. erlernbar - ist, so dass eine höhere Ausprägung des Cues von Entscheidern als Hinweis für eine geringere bzw. höhere Ausprägung des Kriteriums verwendet wird, wenn ein negativer bzw. positiver Zusammenhang zwischen Cue und Kriterium in der Umwelt besteht; siehe auch Unkelbach (2006).

³³Außer eine fixierte Cue-Nutzung führt für spezifische und besonders „fitness“-relevante Situationen zu guten Entscheidungen, die nicht auf Veränderungen in der Umwelt, die möglicherweise nur Rauschen sind, reagiert (Lorenz, 1967; Tinbergen, 1996).

1. Theorie

die Annahme besteht (Goldstein & Gigerenzer, 2002) und es empirische Hinweise dafür gibt (Pachur, Bröder & Marewski, 2008), dass die Rekognitions-Heuristik verstärkt in „natürlichen“ (Goldstein & Gigerenzer, 2002, S. 76) Entscheidungsumwelten, d. h. in Umwelten, die aus realen Stimuli wie den Optionen und Cues in der Städtevergleichsaufgabe bestehen und in denen die Validität des Rekognitions-Cues *vor* einer Studie außerhalb des Labors durch alltägliche Erfahrungen gelernt wurde, angewendet wird und somit auch in natürlichen Umwelten (Brunswik, 1955, 1956; Dhimi et al., 2004) untersucht werden sollte. Es ist dann nicht möglich, die Aufgaben so zu konstruieren, dass die getesteten Modelle unterschiedliche Vorhersagen machen und dadurch erst die Aufgaben informativ für einen Modellvergleich sind.

Die Angabe darüber, wie viele Probanden in wie vielen Vergleichen in die Richtung der Rekognition entscheiden, überschätzt also die Anzahl der Probanden, die die Rekognitions-Heuristik auch tatsächlich verwenden (Hilbig & Richter, 2011). Eine Entscheidung in Richtung des Rekognitions-Cues ist notwendig aber *nicht ausreichend* für den Nachweis der Nutzung der Rekognitions-Heuristik (Hilbig, 2010a, 2010b; Hilbig & Richter, 2011). Hilbig (2010b) demonstriert das anschaulich anhand seiner psychologisch unplausiblen *Alphabet* Heuristik. Ein Anwender der *Alphabet* Heuristik wählt im Vergleich zwischen Objekten das Objekt, dessen Buchstabe im Namen eine höhere Position im Alphabet hat. Ein Anwender der *Alphabet* Heuristik würde z. B. im Vergleich zwischen **B**erlin und **A**achen korrekterweise Berlin wählen, da der Anfangsbuchstabe B nach A im Alphabet kommt und Berlin tatsächlich die größere Stadt ist.

Die *Alphabet* Heuristik zeigt in der Städtevergleichsaufgabe im Vergleich zur Rekognitions-Heuristik und *Fluency* Heuristik hohe Akkuratheitsraten, was für die normative Güte der Heuristik spricht. In einer Reanalyse empirischer Daten zur Städtevergleichsaufgabe (Hilbig, 2010b) zeigte sich, dass Probanden in Richtung der *Alphabet* Heuristik entscheiden. Das spricht nur für die deskriptive Güte der Heuristik, wenn übersehen wird, dass die Bestimmung der Nutzer der Heuristik über ein *inkorrektes* Messmodell erfolgt ist, das nicht sensitiv für die Überschneidung der Modellvorhersagen ist. Es wird hier auch noch einmal deutlich, dass die Bestimmung der normativen Güte einer Heuristik nicht ausreicht, um etwas über die deskriptive Güte bzw. psychologische Plausibilität einer Heuristik auszusagen: Nur weil eine Heuristik adaptiv ist, ist sie noch lange nicht ein gutes Modell menschlichen Entscheidungsverhaltens (Bröder, 2000b).

Hilbig, Erdfelder und Pohl (2010) entwickelten daher ein multinomiales Modell, das *r-model*, in dem die Wahrscheinlichkeiten für kognitive Prozesse, die an einer Entscheidung beteiligt sind, anhand von kategorialen Häufigkeiten über die Objekte, über die ein Proband Rekognition versus keine Rekognition hat, geschätzt werden können. So kann die Ratewahrscheinlichkeit, wenn beide Objekte unbekannt sind, die Wahrscheinlichkeit einer korrekten Entscheidung basierend auf Wissen, wenn beide Objekte bekannt sind, und die Wahrscheinlichkeit der Nutzung der Rekognitions-Heuristik und die Wahrscheinlichkeit der kompensatorischen Nutzung des Rekognitions-Cues mit Wissen-Cues, wenn ein Objekt bekannt ist, geschätzt werden.

Weiterhin kann die Passung unterschiedlicher Modelle auf empirische Daten bestimmt werden, was für den interessierenden Vergleich zwischen der Implementierung der Rekognitions-Heuristik im *r-model* und dem allgemeinen *r-model*, das die nonkompensatorische Nutzung des Rekognitions-Cues beinhaltet, notwendig ist.

Die Befunde zum *r-model* basieren auf 8 Studien mit 400 Probanden und insgesamt 40,000 Entscheidungen für unterschiedliche Inhaltsdomänen von Entscheidungen (Hilbig et al., 2010, S. 131). Die Ergebnisse zeigen, dass die Rate der Nutzung der Rekognitions-Heuristik gemessen über das *r-model* im Mittel über alle Datensätze hinweg bei 63% der Entscheidungen liegt und um circa 20% niedriger als die Schätzung der Anwendung der Rekognitions-Heuristik ausfällt, wenn der methodisch inkorrekte Index, die Rate der Entscheidungen in Richtung der Rekognition, für die Bestimmung des Grads der Nutzung der Rekognitions-Heuristik verwendet wird (Hilbig et al., 2010, Table 2, S. 128). Weiterhin verschlechtern sich die Passungs-Indizes zwischen dem Modell und den empirischen Daten signifikant, wenn die Anwendung der Rekognitions-Heuristik im *r-model* in Höhe der Rate, in der die Probanden in Richtung des Rekognitions-Cues entschieden haben, fixiert wird (Hilbig et al., 2010, Table 2, S. 128).

Es ist jedoch zu beachten, dass die Rekognitions-Heuristik ein Spezialfall des *r-model* ist, somit wird das *r-model* zwangsläufig mindestens genauso gut und durch den zusätzlichen freien Parameter wahrscheinlich besser auf die Daten passen (Cutting, 2000). Hilbig und Richter (2011) konnten jedoch zeigen, dass die Rekognitions-Heuristik im Vergleich zu dem allgemeinen *r-model* schlechter nach dem Bayesianischen Informations-Kriterium, in dem die Anzahl der freien Parameter eingerechnet wird, abschneidet.

Über die Inhaltsdomänen hinweg zeigt sich, dass die Nutzung der

Rekognitions-Heuristik mit der Validität des Rekognitions-Cues positiv korreliert ist (Hilbig et al., 2010, S. 130; siehe auch Pohl, 2006), so dass Hilbig et al. (2010, S. 132) schlussfolgern, dass die Rekognitions-Heuristik nur angewendet wird, wenn der Rekognitions-Cue die einzig valide Information in der Entscheidungsumwelt ist.

1.2.6. Abschließende Bewertung

Neben der in diesem Kapitel eingeführten *Take The Best* Heuristik, der *Take The Last* Heuristik, der *Minimalist* Heuristik, *Tallying*, *Dawes Rule*, *Weighted Tallying*, *Franklin's Rule* und der Rekognitions-Heuristik wurden weitere Heuristiken der adaptiven Werkzeugkiste eines Entscheiders (Gigerenzer et al., 1999) für unterschiedliche Typen von Entscheidungsaufgaben identifiziert³⁴: So gibt es z. B. die *QuickEst* Heuristik für Schätzaufgaben (Hertwig et al., 1999; Martignon & Schmitt, 1999), die *Fluency* Heuristik für Kategorisierungsaufgaben (Schooler & Hertwig, 2005; Hertwig et al., 2008), die *Priority* Heuristik für Lotterier-Präferenzaufgaben (Brandstätter, Gigerenzer & Hertwig, 2006), die *Equity* Heuristik für Ressourcen-Allokationsaufgaben (Hertwig, Davis & Sullo way, 2002) und die *Natural-Mean* Heuristik für Präferenzentscheidungen (Hertwig & Pleskac, 2008).

So lange eine Heuristik so genau spezifiziert ist, dass ihre normative und deskriptive Güte untersucht und somit die Hypothese, dass Entscheider die Heuristik verwenden, falsifiziert werden kann (Glöckner & Betsch, 2011; Popper, 1966), ist das unproblematisch. Dass Entscheider ein Arsenal an Heuristiken für spezifische Entscheidungsprobleme besitzen, ist ja gerade eine zentrale Idee des Forschungsprogramms der adaptiven Werkzeugkiste (Gigerenzer et al., 1999). Problematisch ist jedoch, dass durch das Hinzufügen neuer Heuristiken etwaige empirische Befunde, die gegen das Forschungsprogramm der adaptiven Werkzeugkiste sprechen, konsistent gemacht werden können und somit das Forschungsprogramm an sich nicht falsifiziert werden kann (Hilbig et al., 2010, Fußnote 1, S. 123).

Es ist außerdem für das Forschungsprogramm der adaptiven Werkzeugkiste problematisch, wenn Alternativmodelle die Daten gleich gut beschreiben und die Entscheidungsprozesse gleich gut erklären können

³⁴Identifikation bedeutet, dass die Heuristiken der adaptiven Werkzeugkiste zum Repertoire eines Entscheiders gehören und somit nur „entdeckt“ werden müssen. Problematisch ist dabei, dass der Größe der adaptiven Werkzeugkiste keine Grenze gesetzt ist (Newell & Shanks, 2007, S. 145).

und gleichzeitig allgemeingültiger sind. Allgemeingültigkeit bedeutet dabei *nicht* die Breite der Vorhersage eines Modells von möglichen empirischen Ergebnissen, die noch kongruent mit dem Modell sind (Roberts & Pashler, 2000, Figure 1, S. 359): Ein Modell, das jedes erdenklich mögliche empirische Ergebnis vorhersagen kann und somit nichts verbietet, ist auch nicht falsifizierbar und somit ein schwaches Modell. Gemeint ist die Breite der Anwendbarkeit eines Modells auf verschiedene *Typen* und *Inhaltsdomänen* (Czerlinski et al., 1999) von Entscheidungsaufgaben und die Möglichkeit des Modells verschiedene *Heuristiken* zu modellieren, die im Rahmen der adaptiven Werkzeugkiste notwendig sind, um die intra- und interindividuelle Varianz (siehe vorheriges Kapitel 1.2.5, S. 38) in der Informationssuche, dem Abbruch der Informationssuche und der Integration der Information zu einer Entscheidung (Gigerenzer & Goldstein, 1996) abbilden zu können. Popper (1966) unterscheidet in diesem Zusammenhang zwischen Allgemeinheit und Bestimmtheit (siehe auch Glöckner & Betsch, 2011): Ein gutes Modell hat eine hohen Grad an Allgemeinheit, wenn es auf viele Phänomene angewendet werden kann, und eine hohen Grad an Bestimmtheit, wenn es spezifische Vorhersagen zu empirisch beobachteten Maßen machen kann.

Es stellt sich also die Frage, ob die theoretische Annahme einer Werkzeugkiste, die mit einem Set an Entscheidungs-Heuristiken/Modellen gefüllt ist, *notwendig* ist, um unterschiedliches Entscheidungsverhalten von Probanden für unterschiedliche Inhaltsdomänen von Entscheidungen zu modellieren, oder ob *ein* Modell das Set an Entscheidungs-Modellen abbilden kann (Newell, 2005; Newell & Shanks, 2007) und damit allgemeingültiger ist (Wason, 1960).³⁵

Wenn man ein Modell für alle Entscheidungen annimmt, kann auch das Homunkulus Problem (Newell, 2005, S. 12) vermieden werden, das darin besteht, dass es bei einem vorhandenen Set an Entscheidungs-Heuristiken eine Meta-Heuristik - den Homunkulus - geben muss, der die „richtige“ Heuristik für eine Entscheidungsumwelt aus dem Set von Heuristiken in der Werkzeugkiste auswählt (siehe aber Rieskamp (2008) für eine Lerntheorie zur Selektion von Strategien). Dieser Vorteil wird aber mit (zumindest) einem freien Parameter in den Unimodellen eingekauft, der dafür sorgt, dass das Modell so flexibel ist, dass unter-

³⁵siehe hierzu auch eine vergleichbare Diskussion zu Dual-Process Modellen (Chaiken & Trope, 1999) versus Uni-Process/Single-Route Modellen (Kruglanski & Thompson, 1999) in der Forschung zu Einstellungen (attitudes)

1. Theorie

schiedliches Entscheidungsverhalten modelliert werden kann.³⁶

Im folgenden Kapitel sollen drei Unimodelle exemplarisch vorgestellt werden, bevor das zu testende Unimodell dieser Arbeit - das *Desired Level of Confidence* (DLC) von Hausmann und Läge (2008) - genauer beschrieben wird.

1.3. Unimodelle einer Entscheidung

Die prozessorientierten³⁷ Unimodelle der Entscheidungspsychologie lassen sich in zwei Klassen einteilen: Modelle, in denen eine Entscheidung als *parallel* ablaufender Prozess modelliert wird, und Modelle, in denen eine Entscheidung als *sequentiell* ablaufender Prozess modelliert wird (Newell & Bröder, 2008). In den folgenden Unterkapiteln werden jeweils exemplarisch zwei Vertreter der Unimodelle dieser Modellklassen beschrieben.

1.3.1. Parallele Informationsverarbeitung

Glöckner und Betsch (2008a) modellieren Entscheidungsprozesse über ein *Parallel Constraint Satisfaction* (PCS) Netzwerkmodell (Read, Vanman & Miller, 1997; Rumelhart, McClelland & The PDP Research Group, 1987). In der PCS Netzwerkarchitektur sind die Optionen und die Cues Knoten, die miteinander verknüpft sind (Glöckner & Betsch, 2008a, Figure 1, S. 218). Knoten können eine positive oder negative Aktivierung haben. Wie stark ein Knoten die Aktivierung eines anderen Knoten im PCS Netzwerk verstärkt oder vermindert, wird über die Höhe und das Vorzeichen der Gewichte der Verknüpfungen zwischen den Knoten modelliert. So ist die Cue Validität in den Gewichten eines Quellknotens zu den Cue-Knoten und das Cue-Muster zwischen den Optionen in den Gewichten zwischen den Cue-Knoten und den Options-Knoten abgebildet.

³⁶Bei den im folgenden Kapitel beschriebenen Modellen ist das z. B. die Transformationsfunktion der Validitäten für die Gewichte im *Parallel Constraint Satisfaction* Netzwerkmodell (Glöckner & Betsch, 2008a) und die Höhe der Abbruchschwelle bei den Evidenzakkumulations-Modellen (Lee & Cummins, 2004).

³⁷Neben Modellen, die auch den Anspruch haben die Prozesse, die zu einer Entscheidung führen, zu modellieren, gibt es (sogenannte paramorphe) Modelle, die nur die Entscheidung vorhersagen. Da sich diese Arbeit mit Prozessmodellen beschäftigt, wird diese Modellklasse nicht näher eingeführt, es sei jedoch hier mit der Prospect Theory von Kahneman und Tversky (1979) ein prominentes Beispiel genannt.

Die Informationen werden im PCS Modell *parallel* verarbeitet: Die Information wird gleichzeitig zu einem kohärenten Gesamtbild beziehungsweise - in Anlehnung an die Gestaltpsychologie - zu einer „Gestalt“ (Glöckner & Betsch, 2008a, S. 216) verarbeitet. In einem iterativen Prozess fließt Aktivierung konstant von einem Quellknoten zu den Cues und von den Cues zu den Optionen und dann zurück zu den Cues. Die Verknüpfungen sind also bidirektional. Die parallele Verarbeitung zeigt sich auch in der Software-Implementierung: Bevor die Aktivierung eines Knotens verändert wird, werden die neuen Aktivierungen *aller* Knoten in einer Iteration berechnet. Der iterative Prozess wird abgebrochen, wenn die Veränderungen der Aktivierungen im PCS Netzwerk unter eine gesetzte Schwelle fallen. Die Vorhersage der Wahl einer Option für einen Entscheider wird über die Aktivierung der Optionsknoten, die Konfidenz in eine Entscheidung über die Differenz der Aktivierungen der Optionsknoten und die Entscheidungszeit über die Anzahl der Iterationen aus dem PCS Netzwerkmodell abgeleitet.

Zwar kann das PCS Modell die Heuristiken der adaptiven Werkzeugkiste über die Setzung der Cue Gewichte imitieren: Für MRDM Strategien werden die Gewichte zwischen dem Quellknoten und den Cues kompensatorisch (also mit ähnlich hoher Ausprägung) gesetzt, während für ORDM Strategien die Gewichte non-kompensatorisch (also mit unähnlich hoher Ausprägung) gesetzt werden, damit die Cues entweder gleich oder verschieden stark auf den Entscheidungsprozess einwirken.. Das PCS Modell ist aber nicht als Werkzeug, das die Modellierung aller Heuristiken in nur einem Modell ermöglicht, sondern vielmehr als theoretischer Gegenentwurf zu dem Forschungsprogramm der adaptiven Werkzeugkiste zu verstehen. So ist eine Annahme, dass die Informationsverarbeitung per default *automatisch* und *unbewusst* (Glöckner & Betsch, 2008b; Wyer & Bargh, 1997) verläuft und deliberative Prozesse nur einsetzen, wenn die automatische Verarbeitung der Information im Netzwerk nicht zu einer genügend konsistenten Lösung führt (Glöckner & Betsch, 2008a, Figure 3, S. 223). Damit ist das Argument der Frugalität der ORDM Heuristiken der adaptiven Werkzeugkiste zumindest für den Fall, in dem alle/mehrere Informationen vorliegen, aufgehoben: Viel Information bedeutet nicht unbedingt einen hohen kognitiven Aufwand in der Verrechnung der Information.

Es gibt erste Evidenz, dass das PCS Modell im Vergleich zu den Heuristiken der adaptiven Werkzeugkiste Entscheidungsverhalten besser abbildet: So benötigen z. B. Probanden weniger Zeit und gaben höhere Konfidenzratings für eine Entscheidung ab, wenn das Cue-Muster

1. Theorie

einer Städtepaarvergleichs-Aufgabe eindeutig auf eine Stadt zeigte, d. h., die beiden validesten von drei Cues zeigen auf die Stadt versus nur der valideste Cue zeigt auf eine Stadt (Glöckner & Betsch, 2008a, Figure S. 219). Dieses Entscheidungsverhalten kann für ein solches Cue-Muster aus dem PCS Modell über eine beschleunigte Konsistenzmaximierung und eine erhöhte Differenz der Aktivationen der Optionsknoten abgeleitet werden (Glöckner, 2006; Glöckner & Betsch, 2008a). Dies sind z. B. Befunde, die nicht aus den Heuristiken der adaptiven Werkzeugkiste abgeleitet werden können: Diese sind blind gegenüber dem Grad der Eindeutigkeit eines Cue-Musters im Bezug auf die Entscheidungszeiten. Weiterhin lassen sich durch die bidirektionale Netzwerkarchitektur zusätzliche Hypothesen, wie z. B. die Umbewertung (*coherence shifts*) der Validität von Cues zum Zwecke der Konsistenz einer Entscheidung, ableiten und empirisch testen (Glöckner, Betsch & Schindler, 2009).³⁸

Juslin und Persson (2002) modellieren Entscheidungsprozesse im PROBEX Modell über Gedächtnisprozesse³⁹. Wenn ein Entscheider ein Kriterium eines Objekts, wie z. B. für eine bestimmte Stadt die Einwohnerzahl in der Städtevergleichsaufgabe, einschätzen soll, vergleicht er das Cue-Muster des Objekts mit Cue-Mustern von Objekten, den sogenannten Exemplaren, die aus dem Gedächtnis abgerufen werden und über die Kriteriums-Wissen besteht. Die Modellvorhersage für die Schätzung eines Entscheiders bezüglich des Kriteriumswertes eines vorliegenden Objekts beinhaltet dann den Mittelwert der gewichteten Kriteriumswerte der Exemplare aus dem Gedächtnis (Juslin & Persson, 2002, Formel 6, S. 570). Die Höhe des Gewichts wird über die Ähnlichkeit des Cue-Musters des vorliegenden Objekts mit dem jeweiligen Exemplar bestimmt, so dass ähnliche Exemplare ein hohes versus unähnliche Exemplare ein niedriges Gewicht erhalten und in die Modellvorhersage der Schätzung eines Entscheiders stärker versus schwächer einfließen (Juslin & Persson, 2002, Formel 1, S. 566). Die Anzahl der Exemplare, die für eine Entscheidung abgerufen werden, kann über einen freien Parameter für Probanden geschätzt werden (Juslin & Persson, 2002, S. 572). Konzeptuell kann über den freien Parameter eine ver-

³⁸Siehe weiterhin den Artikel von Jekel, Glöckner, Fiedler und Bröder (unter Begutachtung) für einen normativen Vergleich des PCS Modells mit den Heuristiken, in dem gezeigt wird, dass die Wahlen nach PCS stärker mit den Wahlen der rationalen Lösung (d. h., Wahlen nach Bayes-Theorem) überschneiden.

³⁹Siehe auch die verschiedenen Abwandlungen (Dougherty et al., 1999, 2008) des MINERVA Modells von Hintzman (1988), das Entscheidungsprozesse auch über Gedächtnisprozesse modelliert und ein erfolgreiches Gegenmodell zur Rekognitions-Heuristik ist (Dougherty et al., 2008, S. 207).

änderbare Schwelle des Abbruchs der Informationssuche, d. h. Exemplarsuche, modelliert werden; siehe dazu die Evidenzakkumulations-Modelle im folgenden Kapitel 1.3.2 (S. 56).

Die Aktivierung der Exemplare aus dem Langzeit-Gedächtnis für den Abgleich mit den Objekten einer Entscheidungssituation läuft schnell und parallel ab, der Vergleich des vorliegenden Objekts mit den Exemplaren läuft jedoch sequentiell ab (Juslin & Persson, 2002, S. 569). Das Modell lässt sich also nicht eindeutig den parallelen versus sequentiellen Modellen einer Entscheidung zuordnen; siehe aber Newell und Bröder (2008, Figure 1, S. 200), die PROBEX den parallelen Prozessmodellen einer Entscheidung zuordnen.

Das PROBEX Modell zeigt klare konzeptuelle/theoretische Vorteile gegenüber den Heuristiken der adaptiven Werkzeugkiste. So ist die Annahme nicht notwendig, dass Entscheider *vor* einer Entscheidung die Parameter einer Heuristik - wie die Validitäten oder Diskriminanzraten - abstrahiert repräsentiert haben müssen (Dougherty et al., 2008; Juslin & Persson, 2002; Newell & Shanks, 2007): Der Abgleich eines vorliegenden Objekts mit den Gedächtnisexemplaren läuft *während* des Entscheidungsprozesses ab. Das ist auch deswegen sinnvoll, weil ein Entscheider vor einer Entscheidungsaufgabe nicht wissen kann, welches der Merkmale eines Objekts Cues und welches Merkmal das Kriterium sein wird, so dass nicht für jedes mögliche Kriterium die Validitäten der Cues berechnet und präsent sein müssen (Newell & Bröder, 2008, S. 201).

Juslin und Persson (2002) argumentieren, dass die „Einfachheit“ der Heuristiken über die rechenintensive Komputation der Parameter (Juslin & Persson, 2002, S. 573) erkaufte wird. Todd und Dieckmann (2005) argumentieren jedoch, dass die genauen Validitäten eines Cues zur Anwendung der *Take The Best* Heuristik nicht bekannt sein müssen; wichtig ist nur die Bildung einer korrekten Rangreihenfolge der Cue Validitäten. Eine normativ gute Rangreihenfolge kann dabei über „einfache“ und effektive Regeln im Lernprozess gebildet werden. Todd und Dieckmann (2005) schlagen Lernregeln für Cue Rangreihenfolge vor, die, wie sie in einer Simulationsstudie zeigen, zu guten Akkuratheitsraten in der Städtepaarvergleichs-Aufgabe führen. Theoretisch unbefriedigend ist hier jedoch, dass neben den Regeln der Heuristiken dann auch Regeln für die Bildung der Rangreihenfolge benötigt werden. Das sind dann also Heuristiken für Heuristiken. Ein Modell, das diese Zusatzannahmen nicht benötigt, ist dann womöglich nach dem Kriterium der Einfachheit eines Modells (Grünwald, 2000) theoretisch sinnvoller.

1. Theorie

Persson und Rieskamp (2009) konnten jedoch zeigen, dass PROBEX im Vergleich zu den Heuristiken der adaptiven Werkzeugkiste zu Entscheidungsverhalten von Probanden in einer gedächtnisbasierten Paarvergleichsaufgabe schlechter passt, woraus Persson und Rieskamp (2009, S. 34) schließen, dass Heuristiken der adaptiven Werkzeugkiste Entscheidungen aus dem Gedächtnis besser abbilden. Bröder et al. (2010) konnten jedoch zeigen, dass diese Schlussfolgerung keine Allgemeingültigkeit hat: Das Repräsentationsformat der Cues ist z. B. ein entscheidender Faktor, ob Probanden exemplarbasierte Prozesse oder Heuristiken in ihrer Entscheidung anwenden. Karlsson, Juslin und Olsson (2008) schließen, dass exemplarbasierte Entscheidungen nur in bestimmten Situationen verwendet werden, wie das z. B. der Fall ist, wenn die Umwelten zu komplex werden, da die Cues nonlinear verknüpft sind (Juslin, Karlsson & Olsson, 2008). Das bedeutet also, dass exemplarbasierte „Strategien“ *nur* ein weiteres Werkzeug in der adaptiven Werkzeugkiste eines Entscheiders für gedächtnisbasierte Entscheidungen sind und nicht als Gegenmodell zu den Heuristiken der adaptiven Werkzeugkiste verstanden werden können.

1.3.2. Sequentielle Informationsverarbeitung

Neben den schon besprochenen (Kapitel 1.2, S. 25) Heuristiken der adaptiven Werkzeugkiste (Gigerenzer et al., 1999), in denen Cues sequentiell abgefragt werden, zählen auch Evidenzakkumulations-Modelle (Hausmann & Läge, 2008; Lee & Cummins, 2004) zu den Modellen, in denen Informationen sequentiell verarbeitet werden.

Die Grundidee der Evidenzakkumulations-Modelle ist, dass sich die kompensatorischen versus nonkompensatorischen Strategien auf einem Kontinuum der Informations-Suche anordnen lassen, dessen Extrema die Rate-“Strategie“, also eine Entscheidung basierend auf *keinem* Cue, und Strategien, die *alle* verfügbaren Cues verwenden, sind (Newell, 2005, Figure 1, S. 13).

In der Modellvorstellung der Evidenzakkumulations-Modelle sammelt ein Entscheider so lange Evidenz, bis die Evidenz für eine der Optionen stark genug ist, um über eine Schwelle⁴⁰ zu gelangen (Newell, 2005, S. 13). Die Schwelle kann eine absolute Evidenzschwelle oder eine

⁴⁰Siehe auch *Decision Field Theory* von Busemeyer und Townsend (1993) für ein Schwellenmodell zur Modellierung von Entscheidungsverhalten in Präferenzaufgaben: Die Suche nach Informationen wird abgebrochen, wenn die Präferenz für eine Option eine bestimmte Stärke ausgebildet hat.

relativ zu der Stärke der Evidenz der anderen Optionen festgelegte Evidenzschwelle sein (Hausmann & Läge, 2008, S. 231). Die Schwelle kann für unterschiedliche Entscheider (interindividuelle Varianz) und unterschiedliche Situationen (intraindividuelle Varianz) unterschiedlich hoch ausfallen.

Eine bestimmte Strategie der adaptiven Werkzeugkiste ist demnach nur ein Spezialfall eines generelleren Evidenzakkumulations-Modells. Diese Eigenschaft von Evidenzakkumulations-Modellen wird über (einen) freie(n) Parameter, der/die über ein Set von Aufgaben für Probanden geschätzt werden muss/müssen, *erkauft*, d. h., dass das Modell für seine Flexibilität durch seine(n) freien Parameter in einem Passungsindex wie dem Bayesianischen Informations-Kriterium (Schwarz, 1978) bestraft wird, in dem jeder freie Parameter Punkte auf dem Konto der Passung kostet.⁴¹

Im Gegensatz dazu besitzen die Heuristiken der adaptiven Werkzeugkiste keine freien Parameter: Die Parameter der Heuristiken der adaptiven Werkzeugkiste wie z. B. die Validität, Diskriminanzrate und Successrate sind *keine* freien Parameter, da diese anhand der Lernumwelt berechnet und nicht basierend auf dem Entscheidungsverhalten der Probanden geschätzt werden müssen. Die Annahme, dass die ökologischen Validitäten gleich den subjektiven Validitäten entsprechen, ist jedoch kritisch; siehe dazu auch die Anmerkungen zum Brunswikschen Linsenmodell auf S. 22 in Kapitel 1.1. Die Flexibilität der Werkzeugkiste wird nicht über die Flexibilität der Heuristiken sondern über die Größe des Sets von Heuristiken *erkauft*. Strenggenommen ist hier jedoch „erkauft“ kein angemessener Begriff, da es a) keinen Passungsindex gibt, der die Passung für die Anzahl der Modelle eines gesamten Forschungsprogramms korrigiert, und, selbst wenn es so einen Index geben würde, b) die Anzahl der Heuristiken/Modelle in der adaptiven Werkzeugkiste unbekannt ist. Das erste Problem lässt sich über die Modellierung eines Wahlmechanismus eines Entscheiders für eine bestimmte Heuristik aus der adaptiven Werkzeugkiste als freie(n) Parameter lösen (Rieskamp, 2006, 2008; Rieskamp & Otto, 2006), das zweite Problem ist nicht lösbar und ist letztendlich eine Schwäche des Forschungsprogramms.

⁴¹Eine weitere Methode aus der Informatik für den Modellvergleich ist die Modellevaluation nach dem Prinzip der minimalen Beschreibungslänge („minimum description length“, MDL) (Grünwald, 2000, 2005). Im MDL Ansatz wird die Beschreibungslänge des Modells mit der Beschreibungslänge der Daten, nachdem diese mit dem Modell komprimiert wurden (= durch das Modell erfasste Systematik), verglichen. Ein Modell ist dann gut, wenn es simpel ist (= wenige Parameter) und die Regularitäten in den Daten gut beschreiben kann.

1. Theorie

Zwar schreibt Newell (2005, S. 13), dass es notwendig ist, empirische Techniken zu entwickeln, um die Heuristiken der adaptiven Werkzeugkiste gegen Evidenzakkumulations-Modelle zu testen, er beschreibt aber nicht, wie das konkret möglich sein soll: Evidenzakkumulations-Modelle haben ja gerade die Eigenschaft, die Heuristiken aus der adaptiven Werkzeugkiste in einem Modell *imitieren* zu können, so dass es unmöglich ist, experimentelle Bedingungen zu finden, die die beiden Klassen von Theorien trennen können.⁴² Das bedeutet auch, dass ein Evidenzakkumulations-Modell immer mindestens so gut wie eine Heuristik auf die Daten passen *muss*. Es steht jedoch offen, ob sich aus den Modellen, wenn sie denn nicht über das Maß der Entscheidungen trennbar sind, diskriminierende Vorhersagen auf weiteren Maßen ableiten lassen (z. B. Entscheidungszeiten, Informationssuche, physiologische Erregung).

Es ergeben sich darüber hinaus aus den Evidenzakkumulations-Modellen interessante neue Hypothesen; so stellt sich die Frage, ob die interindividuell sich unterscheidende Schwelle auch eine psychologische Bedeutung hat, d. h., ob die Schwelle eine stabile Persönlichkeitsvariable ist, die mit verwandten Persönlichkeitskonstrukten zusammenhängt, und ob sie mit situationalen Bedingungen interagiert.

Die Schwelle wird dabei in den Evidenzakkumulations-Modellen unterschiedlich modelliert. Im Evidenzakkumulations-Modell von Lee und Cummins (2004) werden die logarithmierten Odds, d. h. die relative Evidenz für die Optionen, mit jedem hinzukommenden Cue nach den Regeln des Bayes-Theorems erneut berechnet (Bayesian updating) (Lee & Cummins, 2004, S. 346). Der Odds Ratio ist die Wahrscheinlichkeit, dass Option A einen höheren Kriteriumswert hat, geteilt durch die Wahrscheinlichkeit, dass Option B einen höheren Kriteriumswert hat, bei gegebenem Cue-Muster. Der Odds Ratio wird logarithmiert, so dass ein Wert größer 0 für Option A, ein Wert kleiner 0 für Option B und ein Wert gleich 0 weder für Option A noch für Option B spricht.

Wenn der logarithmierte Odds Ratio eine Schwelle überschreitet (Lee & Cummins, 2004, Figure 1, S. 346), wird die Suche nach weiteren Cues abgebrochen und in Richtung der Option entschieden, die der logarith-

⁴²siehe dazu auch Newell und Bröder (2008, S. 201): „[...] contingency models [zu denen die Heuristiken der adaptiven Werkzeugkiste gehören] and accumulation models are very hard to distinguish empirically. Whether this matter can be resolved by inventing clever testing methods remains to be investigated [...]. If the issue cannot be addressed empirically, the question may have to be resolved by determining which of the metaphors is more fruitful and fits better into the nomological network of other theories of cognition and decision making.“

mierte Odds Ratio vorgibt. Wenn alle Cues eingeholt wurden und die Evidenz-Schwelle nicht überschritten wurde, wird nach dem Modell in Richtung der Option entschieden, für den der logarithmierte Odds Ratio spricht. Eine individuelle Schwelle kann über ein Set von Aufgaben für individuelle Probanden oder über die Gruppe hinweg geschätzt werden.⁴³ Lee und Cummins (2004) konnten in einer probabilistischen Entscheidungsaufgabe zeigen, dass ihr Evidenzakkumulations-Modell 84.5% der Entscheidungen über die Probanden hinweg korrekt abbilden kann im Gegensatz zu der Akkuratheit von 36% für *Take The Best* und der Akkuratheit von 64% eines Modells, in dem es keine Abbruchschwelle gibt - bzw. die Abbruchschwelle genügend hoch ist - und alle Cues integriert werden. Unverständlicherweise wurden in der Studie alle Cues gleichzeitig präsentiert, so dass die Stopp-Regel der Probanden, über die ja in dem Modell eine Vorhersage ableitbar ist, nicht untersucht wurde.

Es ist nicht klar, ob Lee und Cummins (2004) annehmen, dass Probanden das Bayes-Theorem zur (komplexen) Berechnung der logarithmierten Odds verwenden, was, wenn das der Fall wäre, kritisch für die psychologische Plausibilität des Modells ist. So kann das Evidenzakkumulations-Modell „nur“ in die Klasse der paramorphen Modelle (oder auch as-if Modelle) eingeordnet werden, die den Outcome (Entscheidung) vorhersagen, aber nichts über die psychologischen Prozesse, die zu dem Outcome führen, aussagen (Lewandowsky & Farrell, 2011).

Hausmann und Läge (2008) modellieren eine Schwelle für den Abbruch der Cue-Suche direkt über die Validitäten eines Cues. Wenn ein Cue eine Validität hat, die gleich oder über einer Schwelle liegt, wird die Cue-Suche abgebrochen, wenn das nicht der Fall ist, wird sie fortgeführt.

Die Schwelle ist konzeptuell ein individuelles Anspruchsniveau an *gewünschter* Urteilssicherheit (Hausmann-Thürig, 2004) - das *Desired Level of Confidence* (DLC) (Hausmann & Läge, 2008) -, das von der Evidenz befriedigt werden muss, bevor die Informations-Suche abgebrochen wird.⁴⁴ Es handelt sich nach Hausmann und Läge (2008, Fußnote 1, S. 231) beim DLC um einen emotional-motivationalen Zustand.

Da die Validierung des DLC Modells (Hausmann & Läge, 2008) The-

⁴³Lewandowsky und Farrell (2011) empfehlen prinzipiell eine Modellierung auf Individualebene und Mittelung der individuellen Parameter, falls gruppenstatistische Aussagen getroffen werden sollen.

⁴⁴vergleiche auch mit dem *satisficing* Prinzip von Simon (1990) auf S. 22 in Kapitel 1.1 und mit dem *sufficiency principle* von Chaiken et al. (1989)

1. Theorie

ma der Arbeit ist, wird in den beiden folgenden Unterkapiteln die Aufgabe (Kapitel 1.3.2.1, S. 60) und das zugrundeliegende Messmodell (Kapitel 1.3.2.2, S. 62) der DLC Schwelle beschrieben, bevor die bisherigen empirischen Ergebnisse zum DLC Modell (Kapitel 1.3.2.3, S. 65) genannt werden.

1.3.2.1. Aufgabe Desired Level of Confidence

In der Pferdewette Aufgabe nehmen die Probanden die Rolle eines Wettbüroleiters ein. Ihre Aufgabe ist es, in 71 Durchgängen jeweils 1000 Franken Einsatz auf das Pferd zu setzen, das nach ihrer Meinung in einer simulierten Pferdewette gewinnen wird. Bei einem richtigen Tipp, erhalten die Probanden 300 Franken „Provision“ auf ihr Konto gutgeschrieben; siehe Abbildung 1.4 (S. 61) für die Oberfläche der Pferdewette. Gezeigt wird hier die Oberfläche, die im empirischen Teil verwendet wird. Bis auf die Währung Euro vs. Franken ist der Inhalt und das Design der Aufgabe identisch zu der Studie von Hausmann und Läge (2008, Figure 1, S. 233).

Die vier Pferde haben die gleiche a-priori Wahrscheinlichkeit ($p = .25$) das Siegerpferd zu sein. Die Pferdenamen, Zufallsziehungen aus einem Set klassischer Götternamen, unterscheiden sich über die Durchgänge, so dass den Probanden über die Instruktion hinaus vermittelt wird, dass die Durchgänge voneinander unabhängig sind; siehe für die exakten Instruktionen Anhang B.1 (S. 175).

Die Probanden können, um die Wahrscheinlichkeit eines korrekten Hinweises zu erhöhen, voneinander unabhängige Informanten sequentiell befragen. In jedem Durchgang muss mindestens ein „Hinweis“ von einem Informanten eingeholt werden und es können maximal sieben Informanten befragt werden. Jeder Informant gibt einen Tipp ab, welches Pferd seiner Meinung nach gewinnen wird. Wie gut bzw. wie valide der Hinweis ist, wird über eine weitere Information zum Hinweis vermittelt: Als Güte-Index eines Informanten wird die Anzahl der Fälle angegeben, in denen der Informant in den letzten 100 Pferdewetten das Gewinnerpferd vorhergesagt hat.

Die Befragung der Informanten ist mit 30 Franken Kosten pro Informant verbunden. In der Begrifflichkeit der adaptiven Werkzeugkiste sind die Informanten also die Cues einer probabilistischen Schätzaufgabe, die mit einer bestimmten Validität auf das Kriterium - das Gewinnerpferd - zeigen. Die Diskriminanz liegt aufgrund der Personalisierung der Cues für alle Cues bei 1: Ein Informant gibt immer nur einen Tipp

1.3. Unimodelle einer Entscheidung

Es sind 4 Pferde am Start

Pferd A
Snotra

Pferd B
Freya

Pferd C
Heimdall

Pferd D
Apsara

Aufgabe Nr.

Sie setzen auf Pferd

| Hinweise | Tipp | Anzahl korrekter Vorhersagen in den letzten 100 Pferderennen |
|-----------|------|--|
| Hinweis 1 | B | 74 |
| Hinweis 2 | A | 32 |
| Hinweis 3 | B | 54 |
| Hinweis 4 | | |
| Hinweis 5 | | |
| Hinweis 6 | | |
| Hinweis 7 | | |

Das Aufdecken eines einzelnen Hinweises kostet 30 Euro

nächster Hinweis

aktueller Kontostand (Euro)
910

Es muss mindestens ein Hinweis aufgedeckt werden!

Ihre Provision bei korrekter Vorhersage beträgt 300 Euro.

Abbildung 1.4. Die Oberfläche der Pferdewette. Ein Entscheider muss in einer simulierten Pferdewette zwischen vier Pferden auf das Siegerpferd tippen. Ein Proband mit einem aktuellen Kontostand von 910 Euro Spielgeld hat drei Hinweisreize aufgedeckt, die nur mit einer bestimmten Validität das Gewinnerpferd anzeigen.

für *ein* Pferd ab (Hausmann & Läge, 2008, S. 232).

Nach Abgabe eines Tipps erhalten die Probanden eine Rückmeldung über das tatsächliche Gewinnerpferd, bei einem richtigen Tipp über

1. Theorie

ihren Gewinn in dem Durchgang nach Abzug der Suchkosten für die aufgedeckten Cues bzw. bei einem falschen Tipp über den Verlust in Höhe der investierten Suchkosten für die aufgedeckten Cues und über den aktuellen Kontostand über alle Durchgänge.⁴⁵

1.3.2.2. Messmodell Desired Level of Confidence

Den Probanden werden in der Messphase 51 Durchgänge der Pferdewette präsentiert. Der erste Cue nimmt über die 51 Aufgaben hinweg in randomisierter Reihenfolge alle Validitätswerte zwischen 0 und 1 in Schritten von .02 (0, .02, . . . , .98, 1) an. Die Validitäten der restlichen 6 Cues liegen zufällig zwischen 0 und 1. Zur Berechnung des DLC ist nur die Höhe der Validität des ersten Cues relevant und ob der Proband mehr als nur den ersten Cue in einem Durchgang aufdeckt, bevor er einen Tipp abgibt.

Die Berechnung des DLC lässt sich über die folgende Formel 1.1 darstellen:

$$\begin{aligned} Fehler_{DLC} &= \sum_{i=1}^k \epsilon_i, \\ \epsilon_i &= \begin{cases} (Val_i - DLC)^2 & , \text{ falls MRDM} \mid Val_i \geq DLC \\ (Val_i - DLC)^2 & , \text{ falls ORDM} \mid Val_i < DLC \\ 0 & , \text{ falls MRDM} \mid Val_i < DLC \\ 0 & , \text{ falls ORDM} \mid Val_i \geq DLC \end{cases} \end{aligned} \quad (1.1)$$

Es wird für jedes hypothetische DLC , welches Werte zwischen 0 und 1 in .01 Schritten annehmen kann, über die Summe der Fehler ϵ_i in einem Set von k Aufgaben in der Pferdewette ein Gesamtfehlerterm, $Fehler_{DLC}$, berechnet.

Der Fehler ϵ für die Aufgabe i nimmt die quadrierte Differenz des hypothetischen DLC mit der Validität Val des ersten Cues der Aufgabe i an, wenn die Validität Val_i des ersten Cues *über oder gleich* dem hypothetischen DLC liegt, der Proband jedoch *mehr als einen* Cue in der Aufgabe aufdeckt und somit entgegen der Modellvorhersage MRDM zeigt: Die Validität des Cues liegt über oder gleich der Schwelle und

⁴⁵Die Software funktioniert daher nach folgendem Prinzip: Zuerst wird per Zufall eines der Pferde als Gewinnerpferd gezogen. Dann wird mit einer Wahrscheinlichkeit in Relation zu der Validität eines Cues per Zufallsalgorithmus bestimmt, ob der Cue das Gewinnerpferd anzeigt. Wenn das nicht der Fall ist, wird von den drei Verliererpferden ein Pferd per Zufall für die Cue Anzeige ausgewählt.

sollte dem Anspruchsniveau an gewünschter Urteilssicherheit genügen, trotzdem deckt der Proband weitere Cues im Durchgang auf.

Der Fehler ϵ für die Aufgabe i nimmt die quadrierte Differenz des hypothetischen DLC mit der Validität Val des ersten Cues der Aufgabe i an, wenn die Validität Val_i des ersten Cues *unter* dem hypothetischen DLC liegt, der Proband jedoch nur *einen* Cue in der Aufgabe aufdeckt und somit entgegen der Modellvorhersage ORDM zeigt: Die Validität des Cues liegt unter der Schwelle und sollte dem Anspruchsniveau an gewünschter Urteilssicherheit nicht genügen, trotzdem deckt der Proband nur einen Cue im Durchgang auf. In allen anderen Fällen ist der Fehler ϵ_i gleich dem Wert 0, da sich der Proband unter der Annahme der Gültigkeit des hypothetischen DLC in Richtung der Modellvorhersage verhält. Das hypothetische DLC mit dem geringsten Gesamtfehler $Fehler_{DLC}$ ist der gemessene DLC Schwellenwert für einen Probanden. Das Messmodell bildet also in der jetzigen Spezifizierung keinen Evidenz-*Akkumulations*-Prozess ab, da die Integration der Cues zum einem Gesamt-Sicherheitsindex nicht im Messmodell erfasst wird.

Den Probanden wird in der Testphase, in der das gemessene DLC adaptiv getestet wird, weitere 20 Durchgänge der Pferdewette präsentiert. Dazu werden in zufälliger Reihenfolge 20 weitere Durchgänge der Pferdewette präsentiert, von denen die Validität des ersten Cues in 10 Durchgängen über dem DLC liegt, nämlich in den Abständen $DLC +.03, +.05, \dots, +.19, +.21$, und in 10 Durchgängen unter dem DLC liegt, nämlich in den Abständen $DLC -.03, -.05, \dots, -.19, -.21$. Da sich die ersten 51 Durchgänge der Messphase und die 20 Durchgänge der Testphase oberflächlich nicht unterscheiden und im direkten Anschluss stattfinden, ist dem Probanden nicht bewusst, dass die Pferdewette aus 2 Phasen besteht. In Anhang A.2 (S. 171) sind Veränderungen zur Original Pferdewette und Details zum Algorithmus zur Berechnung der Validitäten des ersten Cues über die 20 Testdurchgänge für die Studien dieser Arbeit beschrieben.

In Tabelle 1.3 (S. 64) ist die erwartete Verteilung der Häufigkeiten der Anwendung einer ORDM Stoppregel und einer MRDM Stoppregel der Cue-Suche über das Set von 20 Testdurchgängen in der Pferdewette für vier unterschiedliche Modelle einer Entscheidung dargestellt.

Wenn das DLC eine valide Messung des Abbruchkriteriums der Cue-Suche ist, dann sollten Probanden in den 10 Durchgängen, in denen die Validität des ersten Cues gleich oder über dem DLC liegt, nur einen Cue aufdecken und somit eine ORDM Stoppregel der Cue-Suche anwenden. In den 10 Durchgängen, in denen die Validität des ersten Cues unter

1. Theorie

Tabelle 1.3. Erwartete Häufigkeiten der Anwendung einer ORDM und einer MRDM Stoppregel der Cue-Suche bei gegebener Validität des ersten Cues eines Testdurchgangs und der gemessenen DLC Schwelle eines Probanden über das Set von 20 Testdurchgängen der Pferdewette für vier konkurrierende Modelle einer Entscheidung: Das DLC Modell, ein ORDM Modell, ein MRDM Modell und ein Zufallsmodell; vergleiche mit Hausmann und Läge (2008, Figure 2, S. 235).

| Stoppregel | Modellvorhersage | | | |
|--------------------------------|------------------|------|------|--------|
| | DLC | ORDM | MRDM | Zufall |
| MRDM $Val_i \geq \text{DLC}$ | 0 | 0 | 10 | 5 |
| ORDM $Val_i < \text{DLC}$ | 0 | 10 | 0 | 5 |
| MRDM $Val_i < \text{DLC}$ | 10 | 0 | 10 | 5 |
| ORDM $Val_i \geq \text{DLC}$ | 10 | 10 | 0 | 5 |

dem DLC liegt, sollte ein Entscheider nach dem DLC Modell mehr als einen Cue aufdecken und somit eine MRDM Stoppregel der Cue-Suche anwenden.

Wenn ein Entscheider einer reinen ORDM Strategie folgt, dann wird er unabhängig vom DLC und der Validität des ersten Cues für alle 20 Testdurchgänge immer nur einen Cue aufdecken. Wenn ein Entscheider einer reinen MRDM Strategie folgt, dann wird er unabhängig vom DLC und der Validität des ersten Cues für alle 20 Testdurchgänge immer mehr als einen Cue aufdecken. Wenn ein Entscheider zufällig eine ORDM oder MRDM Stoppregel der Cue-Suche anwendet, wird er unabhängig vom DLC gleich oft einen oder mehrere Cues aufdecken.

Probanden können individuell einem der Modelle des Abbruchs der Cue Suche zugeordnet werden. Die Anzahl der Durchgänge, in denen modellkonsistentes Stopp-Verhalten gezeigt wird, kann über einen Binomialtest gegen den Zufall abgesichert werden: Bei 20 Durchgängen muss bei einem α -Fehler von .05 und einem p von .5 mindestens in 15 Durchgängen modellkonsistentes Stopp-Verhalten gezeigt werden, um die überzufällige korrekte Vorhersage des Modells für einen Probanden zu sichern.

1.3.2.3. Befunde Desired Level of Confidence

In einer Validierungsstudie zum DLC von Hausmann und Läge (2008, S. 234) zeigte sich, dass das DLC der Probanden zwischen .52 und .88 mit einem Mittelwert von .73 und einer Streuung von .09 lag. Der Gesamtfehlerwert, der mit dem gemessenen DLC verknüpft war, lag zwischen 0 und .11 bei einem Mittelwert von .03 und einer Streuung von .03.

Über das gesamte Set an beobachteten Entscheidungen, das sind also bei 22 Probanden und 20 adaptiven Testdurchgängen 440 Entscheidungen, waren nur in 49% der Entscheidungen die beobachtete Stoppregel der Probanden konform mit einem MRDM Modell und damit in 51% der Entscheidungen konform mit einem ORDM Modell (siehe Tabelle 1.3, S. 64). Dagegen passte in 91% der Entscheidungen die beobachteten Stoppregel zu dem DLC Modell (Hausmann & Läge, 2008, S. 234).

In einem Binomialtest zeigte sich, dass die Anwendung der Stoppregel bei allen - bis auf einen von 22 Probanden - signifikant von einem ORDM-, MRDM- und Zufalls-Modell abwich (Hausmann & Läge, 2008, S. 234). Im Gegensatz dazu gab es keine signifikante Abweichung zwischen dem DLC Modell und der Anwendung der Stoppregel für alle Probanden - bis auf den schon erwähnten Probanden, dessen Anwendung der Stoppregel nicht signifikant von einem MRDM Modell abwich. Das bedeutet also, dass das DLC Modell bei 21 von 22, das sind $\approx 95\%$ der Probanden, am besten zu der beobachteten Anwendung der Stoppregel passte.

In einer weiteren Studie (Hausmann & Läge, 2008) wurden intraindividuell die Merkmale der Umwelt manipuliert: Die Cues wurden über 30 Durchgänge der Pferdewette nach der Validität der Cues angeordnet und die Probanden darüber instruiert, dass der erste Cue auch der valideste Cue ist, und über weitere 30 Durchgänge per Zufall geordnet, was der Standardbedingung der DLC Aufgabe entspricht (siehe Kapitel 1.3.2.1, S. 60). Die Reihenfolge der Umwelten wurde per Zufall bestimmt. Eine Ordnung nach Validitäten ist eine „freundliche“ (Hausmann & Läge, 2008, S. 235) Umwelt für ein ORDM Heuristik wie *Take The Best*, in der die Cues nach der Höhe der Validitäten abgesucht werden, und sollte ihre Anwendung begünstigen.

Generell zeigten die Probanden jedoch mehr MRDM als ORDM Verhalten, selbst in der *Take The Best* freundlichen Umwelt lag die Anwendung einer MRDM Stoppregel bei 63.4% der Durchgänge (Hausmann

1. Theorie

& Läge, 2008, S. 235). Im Vergleich zu der Zufallsreihenfolge, die einer *Minimalist* Heuristik⁴⁶ entgegenkommt, zeigte sich jedoch ein Anstieg, im Mittel von ≈ 8 auf ≈ 10 , in der Anzahl der Durchgänge, in denen die Probanden ausschließlich einen Cue konsultierten, so dass die Probanden trotz genereller Tendenz zu einer MRDM Stoppregel adaptiv auf die Umwelt reagierten.

Das DLC wurde für die beiden Umwelt-Bedingungen separat berechnet. Es zeigte sich keine signifikante intraindividuelle Differenz in der Höhe des DLC. In der Bedingung, in der die Cues zufällig angeordnet waren, passte für 31 von 36 Probanden das DLC Modell im Vergleich zu einem ORDM-, MRDM- und Zufallsmodell am besten auf die Stoppregel der Probanden. In der Bedingung, in der die Cues nach ihrer Validität angeordnet waren, passte bei 26 Probanden das DLC Modell im Vergleich zu einem ORDM-, MRDM- und Zufallsmodell am besten auf die Stoppregel der Probanden. Das DLC zeigt also auch in einer Umwelt, die ORDM Heuristiken begünstigt, eine gute Passung zur Stoppregel der Probanden, was für die Robustheit des DLC Modells spricht.

1.4. Überblick empirischer Teil der Arbeit

Da es zum DLC Modell bisher nur eine publizierte Studie gibt (Hausmann & Läge, 2008), ist eine weitere Validierung des DLC Modells notwendig. Im ersten Abschnitt der Arbeit wird die Reliabilität und Validität des DLC in zwei Studien untersucht.

In Studie 1 des empirischen Teils der Arbeit (Kapitel 2.1, S. 69) wird auf verschiedene Arten die Reliabilität der Messaufgabe des DLC bestimmt. Es wird untersucht, inwieweit das DLC von den messirrelevanten Merkmalen der Aufgabe beeinflusst wird. Dafür wurden alle *freien* Parameter der Pferdewette so verändert, dass die neue Aufgabe - es handelte sich um einen Kriminalfall als Coverstory - möglichst verschieden zur Pferdewette ist. Freie Parameter sind solche Parameter, die verändert werden können, ohne dass sich das Messmodell des DLC verändert: Das ist also die Anzahl der Optionen und Cues, die Cue Kosten, die Höhe des Gewinns bei der richtigen Wahl, die Anordnung der Elemente der Oberfläche und die Coverstory. Es wurde separat für die Pferdewette und den Kriminalfall für jeden Probanden das DLC gemessen und untersucht, wie die DLC Maße zusammenhängen. Weiterhin wird ana-

⁴⁶siehe Seite 29 in Kapitel 1.2.2 für die Beschreibung der *Minimalist* Heuristik

log zu der Originalstudie von Hausmann und Läge (2008) das DLC der Testphase mit dem DLC der Messphase verglichen. Das bedeutet, dass in den 51 ersten Durchgängen das DLC für jede der beiden Einkleidungen der DLC-Aufgabe (Pferdewette und Krimi) ermittelt wird und in den anschließenden 20 Durchgängen (adaptiv) getestet wird, ob das DLC Modell im Vergleich zu ORDM oder MRDM Strategien die Wahlen der Probanden am besten vorhersagen kann (siehe Kapitel 1.3.2.2, S. 62 für Details zum Messmodell). Über die Originalstudie hinaus wird für die 20 Aufgaben der Testphase ein DLC nach dem gleichen Algorithmus wie in der Testphase berechnet, so dass die beiden DLC Maße direkt miteinander korreliert werden können und somit – neben der Vorhersagekraft des DLC Modells – ein weiteres Maß für die Reliabilität des DLC ermittelt werden kann.

In Studie 2 (Kapitel 2.2, S. 79) wird die Konstruktvalidität des DLC bestimmt. Dazu wird untersucht, wie das DLC Modell in das Forschungsprogramm der adaptiven Werkzeugkiste (Gigerenzer et al., 1999) passt. Dabei wird in der ersten Phase der Studie das DLC in der Pferdewette gemessen und in der zweiten Phase das Entscheidungsverhalten der Probanden im Börsenspiel nach ORDM und MRDM Strategien klassifiziert. In der gleichen Studie wurde die Pferdewette und das Börsenspiel eine Woche später noch einmal an den gleichen Probanden erhoben. Damit ist es möglich zu untersuchen, ob es sich beim DLC um ein zeitstabiles Maß handelt, und es sich somit um eine Persönlichkeitseigenschaft handelt, was von Newell und Bröder (2008, S. 198, Hervorhebung hinzugefügt) schon als Hypothese aufgestellt wurde: „Their work [gemeint ist das hier beschriebene DLC Modell von Hausmann und Läge (2008)] is especially interesting because they develop a simple method of validating the threshold concept empirically and at the same time model (*stable?*) *individual differences* under the umbrella of a unified process model.“

In den verbleibenden Studien des empirischen Teils der Arbeit wird untersucht, wie das DLC mit verwandten Konstrukten aus der kognitiven Psychologie *außerhalb* des Forschungsprogramms der adaptiven Werkzeugkiste (Gigerenzer et al., 1999) zusammenhängt.

Dazu werden im zweiten Abschnitt der Arbeit in vier Studien situationale Einflussfaktoren auf das DLC untersucht.

In Studie 3 (Kapitel 3.1.1, S. 93) und Studie 4 (Kapitel 3.1.2, S. 100) wird untersucht, ob das DLC durch ein konzeptuelles Priming oder ein *Mindset* Priming (Bargh & Chartrand, 2000; Uleman & Bargh, 1989) einer ORDM versus MRDM Strategie beeinflusst werden kann. Da in

1. Theorie

Studie 2 untersucht wird, ob das DLC eine stabile Persönlichkeitseigenschaft ist, kann die Folgerung gezogen werden, dass eine situationale Manipulation keinen Einfluss auf das DLC haben *darf*, wenn sich zeigt, dass das DLC zeitlich stabil ist. Die Hypothese bei den experimentellen Studien dieser Arbeit ist jedoch, dass das DLC mit situationalen Einflüssen interagiert, so dass sich der Mittelwert des DLC in den experimentellen Gruppen unterscheiden kann, relative Unterschiede zwischen den Probanden in den Gruppen aber konstante und bedeutungsvolle Unterschiede sind. Es besteht also die Hypothese, dass die Situation additiv auf das DLC wirkt.

In Studie 5 (Kapitel 3.2.1, S. 109) und Studie 6 (Kapitel 3.2.2, S. 113) wird untersucht, ob sich das DLC durch die Induktion eines hohen versus niedrigen *Need for Cognitive Closure* (Kruglanski, 2004; Kruglanski & Webster, 1996) über die Attraktivität der Aufgabe und Lärm in der Versuchsbedingung beeinflussen lässt.

Im abschließenden dritten Abschnitt der Arbeit wird der Zusammenhang zwischen dem DLC und Persönlichkeitsvariablen untersucht.

In Studie 7 (Kapitel 4.1.1, S. 124) und Studie 8 (Kapitel 4.1.2, S. 130) wird untersucht, wie das DLC mit Antworttendenzen zusammenhängt, d. h., ob das DLC mit dem Entscheidungskriterium β der Signalentdeckungstheorie (Macmillan & Creelman, 2005; Stanislaw & Todorov, 1999) in einer Kategorisierungsaufgabe und in einer Rekognitionsaufgabe korreliert ist.

In Kapitel 4.2 (S. 135) wird ein Überblick über den Zusammenhang zwischen dem DLC und Persönlichkeits-Maßen, die über die Studien hinweg erhoben wurden, gegeben. Die Arbeit ist also so organisiert, dass in den einzelnen Methodenteilen zu den Studien die Persönlichkeitstests, die erhoben wurden, genannt werden, die Ergebnisse jedoch erst im Überblick am Ende der Arbeit zusammenfassend dargestellt werden.

Im abschließenden Kapitel 5 (S. 141) der Arbeit werden die Befunde aller Studien zusammengefasst und diskutiert.

2. Reliabilität und Validität

In den folgenden beiden Studien wird die Reliabilität und Validität des DLC untersucht. In der ersten Studie wird die Reliabilität, d. h. die Messgenauigkeit des DLC in der Pferdewettenaufgabe, überprüft. Eine genügend hohe Reliabilität soll garantieren, dass das DLC mit der Messaufgabe genügend zuverlässig gemessen werden kann. In der zweiten Studie soll überprüft werden, ob es sich bei dem DLC um ein zeitstabiles und somit eine Persönlichkeitseigenschaft handelt und ob es stimmig in den Kontext der im Theorieteil dargestellten Befunde der adaptiven Werkzeugkiste einzuordnen ist. In beiden Studien wird also die Grundlage für die weiteren Studien dieser Arbeit gelegt. Wenn das DLC nicht über die Messaufgabe genügend reliabel erfasst werden kann, sind die folgenden Studien, bei denen das DLC situational manipuliert werden soll, obsolet bzw. es ist in einem ersten Schritt notwendig ein reliableres Messverfahren zu entwickeln. Wenn das DLC weiterhin zeitlich instabil ist - es sich also um einen zeitinstabilen Zustand (= State) handelt - , sind Untersuchungen zu den Zusammenhängen zwischen dem DLC und anderen Persönlichkeitsvariablen wenig sinnvoll. Wenn das DLC nicht mit bewährten Maßen der Entscheidungspsychologie zusammenhängt, ist die Frage zu stellen, ob das DLC überhaupt Entscheidungsverhalten außerhalb der Messaufgabe abbilden kann, womit die Konstruktvalidität des DLC in Frage gestellt werden kann.

2.1. Reliabilität des DLC

In der folgenden Studie soll untersucht werden, wie reliabel die Messaufgabe der Pferdewette das DLC erfassen kann. Dazu wird analog zur Studie von Hausmann und Läge (2008) das DLC in einer ersten Phase gemessen und in einer zweiten Phase (adaptiv) getestet, ob das gemessene DLC Entscheidungsverhalten in der Testphase besser als eine ORDM oder MRDM Strategie erklären kann. Es wird somit die prädiktive Kraft des DLC Modells getestet. Darüber hinaus wird über den gleichen Verrechnungsalgorithmus in der Testphase ein DLC berechnet,

2. Reliabilität und Validität

um über die Interkorrelation der beiden DLC Werte ein weiteres Maß der Messgenauigkeit abzuleiten.

In einem zweiten Schritt soll untersucht werden, ob das DLC von den Ausprägungen der Parameter der Messaufgabe beeinflusst wird, die *nicht* relevant für die Anwendung des Messmodells des DLC sind. Um die Stabilität der interindividuellen Unterschiede des DLC zu untersuchen, wurde daher nach der Pferdewette eine zweite Aufgabe, die als Coverstory einen Kriminalfall beinhaltete, erhoben und das DLC der Pferdewette mit dem DLC des Kriminalfalls verglichen. Im Kriminalfall wurden *alle* freien Parameter der Pferdewette verändert, um die Aufgabe größtmöglich unterschiedlich zur Pferdewette zu gestalten. Ein Befund, der z. B. zeigt, dass durch die Anordnung der Elemente auf der Oberfläche die interindividuellen Unterschiede der Probanden auf dem DLC beeinflusst werden, würde entweder a) gegen die Güte der Messaufgabe zur Bestimmung eines individuellen DLC oder b) gegen die Stabilität des DLC gegenüber der Messaufgabe sprechen. Die Analyse dient also dazu, den Einfluss der (irrelevanten) Einkleidung der Messaufgabe auf den DLC Messwert zu untersuchen. Anders formuliert: Die Varianz im Messwert des DLC besteht aus einem Anteil wahrer Konstruktvarianz (das zu messende DLC Konstrukt) und irrelevanter Varianz. Ein Teil der irrelevanten Varianz ist auf die Operationalisierung des DLC in der Messaufgabe zurückzuführen. Wenn man nun die DLC Messaufgabe stark verändert und trotzdem eine hohe Korrelation zwischen den beiden Operationalisierungen des DLC besteht, dann ist dies Evidenz dafür, dass die Messung nicht von irrelevanten Anteilen in der Messaufgabe getrieben wird.

2.1.1. Methoden

Die Studie gliedert sich in vier Phasen. Zuerst wurde das DLC der Pferdewette erhoben, dann das DLC der Kriminalaufgabe, dann weitere Aufgaben, die konstrukt sensitiv für das gewünschte Anspuchsniveau an Urteilssicherheit sein sollen und daher mit dem DLC verglichen werden, und abschließend mit der Erhebung relevanter Persönlichkeitsvariablen, die in Kapitel 4.2 (S. 135) zusammenfassend dargestellt werden. Im Folgenden werden die Aufgaben der einzelnen Phasen genauer beschrieben.

Die Probanden wurden schriftlich über ihre Rechte als Teilnehmer an einer psychologischen Studie aufgeklärt. Die Durchführung aller Aufgaben fand am Computer statt. Die Probanden erhielten für ihre Teil-

nahme für ihr Studium notwendige Teilnahmestunden gutgeschrieben und die drei Probanden mit der höchsten Punktzahl (s. u.) erhielten zusätzlich einen Kinogutschein.

Die Probanden bearbeiteten zuerst die 71 Durchgänge der Pferdewette, die in Kapitel 1.3.2.1 (S. 60) und Anhang A.2 (S. 171) beschrieben ist.

Dann bearbeiteten die Probanden die 71 Durchgänge der Kriminalaufgabe; siehe Abbildung 2.1 (S. 71) für die Oberfläche der Kriminalaufgabe.¹

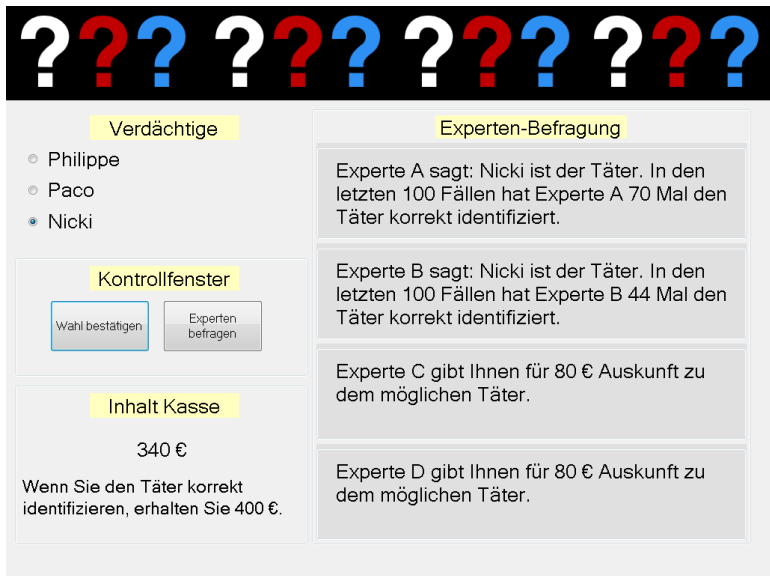


Abbildung 2.1. Die Oberfläche des Kriminalfalls. Ein Entscheider muss in einem simulierten Kriminalfall von 3 Verdächtigen den Täter identifizieren. Ein Proband mit einem aktuellen Kontostand von 340 Euro Spielgeld hat 2 Hinweisreize aufgedeckt, die nur mit einer bestimmten Validität den korrekten Täter anzeigen.

In der Kriminalaufgabe hatten die Probanden die Aufgabe, aus 3

¹Die Coverstory und das Design der Oberfläche der Kriminalaufgabe stammen von Anne Berger, André Bokelmann, Özlem Kara, Zuzana Matulikova und Constanze Rohrmeier und wurden im Rahmen eines Experimentalpraktikums an der Universität Bonn angefertigt.

2. Reliabilität und Validität

Verdächtigen, den Täter in einem *Kriminalfall* zu identifizieren.² Die Namen der Verdächtigen wurden per Zufall ohne Zurücklegen aus einem Set von männlichen Vornamen gezogen, so dass kein Name über die 71 Durchgänge des Kriminalfalls doppelt vorkam. Dazu konnten sie Hinweise von 4 unabhängigen Experten einholen, die mit einer bestimmten Validität den Täter anzeigten. Ein Hinweis eines Experten kostete 80 Euro. Bei einer korrekten Identifikation des Täters erhielten die Probanden 400 Euro Spielgeld Belohnung. Die *Anordnung der Elemente* - wie z. B. die Position der Buttons und die Art der Auswahl eines Verdächtigen - wurde im Vergleich zur Pferdewette verändert.

Die Kriminalaufgabe ist sonst identisch zu der Pferdewette, d. h., dass über 51 Durchgänge hinweg alle Werte zwischen .5 und 1 für die Validitäten des ersten Cues eingespielt wurden und in den abschließenden 20 Aufgaben das DLC der Kriminalaufgabe adaptiv getestet wurde.³

Um zu untersuchen, ob das DLC mit Verhalten (bzw. Verhaltensintentionen, da nur per Fragebogen erhoben) korreliert ist, bearbeiteten die Probanden 13 fiktive Entscheidungssituationen, in denen sie angeben sollten, wie sicher das Eintreten der positiven Konsequenz einer Entscheidungsoption sein muss, damit sie sich für die Option entscheiden. So sollten die Probanden in einem Szenario entscheiden, wie sicher es sein muss, dass eine Therapie gegen Heuschnupfen auch wirklich wirkt, damit sie die Therapie, die mit Risiken wie Nebenwirkungen verbunden ist, auch durchführen lassen; siehe Anhang C.1 (S. 183) für eine Auflistung aller Szenarien.

Die Szenarien stammen aus verschiedenen Lebensbereichen und sind unterschiedlich gravierend in ihren Konsequenzen, so dass Probanden neben dem Sicherheitsurteil auch angeben sollten, wie wichtig ihnen eine solche Entscheidung im realen Leben wäre.

Die Reihenfolge der Entscheidungssituationen wurde randomisiert.

Um zu untersuchen, ob das DLC mit Aufgaben zusammenhängt, die die gewünschte Sicherheit in eine Information abbilden (sollen), bearbeiteten die Probanden in 40 Durchgängen Bilderrätsel, in denen Alltagsgegenstände - wie z. B. eine Uhr, ein Knopf oder eine Tasse - von einer Maske aus 100 Elementen verdeckt waren; siehe Abbildung 2.2 (S. 73).

Mit Drücken des Start Buttons wurde alle 400 Millisekunden per Zufall ein Maskenelement aufgedeckt. Das Aufdecken eines Maskenele-

²siehe Anhang B.3 (S. 177) für die Instruktionen der Kriminalaufgabe

³siehe Kapitel 1.3.2.2 (S. 62) für das Messmodell des DLC

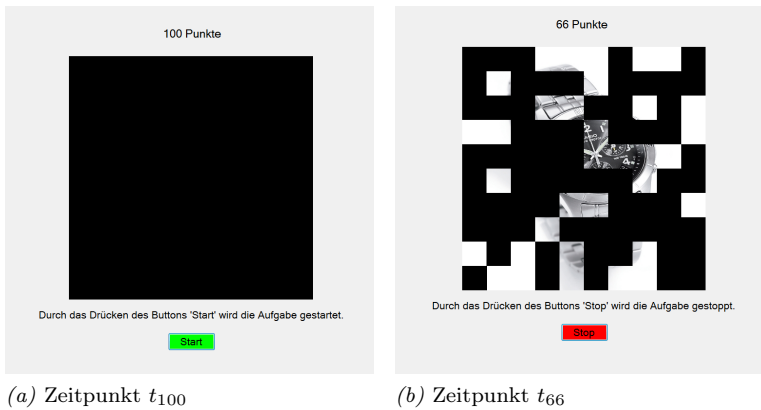


Abbildung 2.2. Die Oberfläche des Bilderrätsels zu zwei Zeitpunkten. Ein Objekt - eine Uhr - befindet sich unter einer Maske aus 100 schwarzen Kästchen. Alle 400 Millisekunden wird 1 Kästchen per Zufall aufgedeckt, was 1 Punkt Spielwährung kostet. Die Probanden sollen erraten, welches Objekt sich unter der Maske befindet, sobald sie sicher genug sind, das Objekt identifiziert zu haben.

ments kostete einen Punkt, so dass die Maximalzahl von 100 Punkten pro aufgedecktem Maskenelement um einen Punkt reduziert wurde und ein Durchgang maximal 40.4 Sekunden dauerte. Das Aufdecken der Maskenelemente konnte durch Drücken eines Buttons gestoppt werden und in einem Folgebildschirm ein Tipp, um welchen Gegenstand es sich handelt, in ein freies Textfeld eingegeben werden. Die Aufgabe der Probanden war es, die maximale Punktzahl bei einem korrekten Tipp zu erzielen.

Die Reihenfolge der Bilderrätsel wurde randomisiert.

Dann wurden das *Bedürfnis nach kognitiver Geschlossenheit* (Schlink & Walther, 2007), der *regulatorische Fokus* (Werth & Förster, 2007) und die *Achievement Motives Scale* (Lang & Fries, 2006) am Computer erhoben; siehe Kapitel 4.2 (S. 135) für die Beschreibung der Skalen, Hypothesen des Zusammenhangs mit dem DLC und den Ergebnissen.

In der letzten Phase der Studie wurden demographische Daten der Probanden erhoben.

2. Reliabilität und Validität

2.1.1.1. Stichprobe

Es nahmen 51 Probanden an der Studie teil. Davon waren 17 männlich und 34 weiblich. Das Alter lag zwischen 19 und 64 Jahren bei einem Mittelwert von 26.8 Jahren und einer Streuung von 10.8 Jahren.

2.1.2. Hypothesen

Das in der Messphase ermittelte DLC kann Entscheidungsverhalten in der Testphase im Vergleich zu einer ORDM und MRDM Strategie am besten vorhersagen. Weiterhin ist das in der Messphase ermittelte DLC positiv mit dem DLC der Testphase korreliert.

Das DLC ist unabhängig von den Merkmalen der Messaufgabe. Es wird eine positive Korrelation zwischen dem DLC der Pferdewette und dem DLC des Kriminalfalls erwartet.

Probanden mit einem hohen DLC geben ein höheres Sicherheits-Rating für das Eintreten einer Konsequenz für eine Entscheidungsoption in den Entscheidungssituationen an. Es wird eine positive Korrelation zwischen dem DLC und den Sicherheits-Ratings erwartet.

Probanden mit einem höheren DLC benötigen mehr Evidenz im Bilderrätsel, bevor sie einen Tipp abgeben. Es wird eine negative Korrelation zwischen dem DLC der Pferdewette und der Punktzahl des Bilderrätsels erwartet.

2.1.3. Ergebnisse

Das mittlere DLC der Pferdewette liegt bei .765 bei einer Streuung von .115. Das mittlere DLC des Kriminalfalls liegt bei .725 bei einer Streuung von .092. Die Modelltestung gegen den Zufall⁴ ergibt, dass 34 von 51 (66%) Probanden als Anwender einer Stoppregel der Informationssuche, die mit dem Pferdewette DLC konform ist, in den 20 Testdurchgängen der Pferdewette klassifiziert werden können. 6 Probanden können als Anwender einer ORDM-, 4 Probanden als Anwender einer MRDM- und 2 Probanden als Anwender einer Zufalls-Stoppregel klassifiziert werden. 5 Probanden lassen sich nicht eindeutig nur einem der Modelle zuordnen.

Die Modelltestung gegen den Zufall ergibt, dass 45 von 51 (88%) Probanden als Anwender einer Stoppregel der Informationssuche, die mit dem Kriminalfall DLC konform ist, in den 20 Testdurchgängen

⁴siehe S. 64 in Kapitel 1.3.2.2 für Details zur Testung gegen den Zufall

des Kriminalfalls klassifiziert werden können. 2 Probanden können als Anwender einer ORDM- und 2 Probanden als Anwender einer Zufalls-Stoppregel klassifiziert werden. 2 Probanden lassen sich nicht eindeutig nur einem der Modelle zuordnen.

In Tabelle 2.1 sind die Häufigkeiten der Modellklassifikation für das DLC der Pferdewette und des Kriminalfalls gegeneinander in einer Kreuztabelle abgetragen.

Tabelle 2.1. Kreuztabelle der Modellklassifikation über die 20 Testdurchgänge des DLC der Pferdewette und der Kriminalaufgabe.

| Modell- klassifikation Pferde | Modellklassifikation Krimi | | | | | Σ |
|-------------------------------------|----------------------------|------|------|--------|-------------|----------|
| | DLC | ORDM | MRDM | Zufall | Uneindeutig | |
| DLC | 32 | 0 | 0 | 2 | 0 | 34 |
| ORDM | 3 | 2 | 0 | 0 | 1 | 6 |
| MRDM | 4 | 0 | 0 | 0 | 0 | 4 |
| Zufall | 2 | 0 | 0 | 0 | 0 | 2 |
| Uneindeutig | 4 | 0 | 0 | 0 | 1 | 5 |
| Σ | 45 | 2 | 0 | 2 | 2 | (51) |

Die Korrelation zwischen dem DLC der Messphase und der Testphase in der Pferdewette liegt bei einem Pearson's $r = .84$, $t(49) = 11.01$, $p < .05$. Die Korrelation zwischen dem DLC der Messphase und der Testphase in dem Kriminalfall liegt bei einem Pearson's $r = .88$, $t(49) = 13.65$, $p < .05$. Das DLC der Pferdewette (und des Kriminalfalls) ist somit befriedigend reliabel.

Die Hypothese über die positive Korrelation zwischen dem DLC der Pferdewette und dem DLC des Kriminalfalls kann bestätigt werden⁵. Die Probanden, die ein höheres DLC in der Pferdewette haben, haben auch ein signifikant höheres DLC im Kriminalfall. Die Korrelation liegt bei Pearson's $r = .79$, $t(49) = 9.2$, $p < .05$.

Die Hypothese über die positive Korrelation zwischen dem DLC der Pferdewette und den abgegebenen Sicherheits-Ratings der Entscheidungssituationen kann nicht bestätigt werden. Es ist nicht der Fall, dass Pro-

⁵Die Power liegt bei einer erwarteten hohen Effektstärke von Pearson's $|r| = .5$ und einer Stichprobengröße von 51 und einem $\alpha = .05$ bei $1 - \beta = .99$.

2. Reliabilität und Validität

banden mit einem höheren DLC auch signifikant mehr Sicherheit für die Wahrscheinlichkeit des Auftretens einer positiven Konsequenz einer Entscheidungsoption benötigen.

In 2 von 13 Entscheidungssituationen korreliert das DLC der Pferdewette hypothesen*inkonform* signifikant negativ mit dem Sicherheits-Rating; siehe Tabelle 2.2 (S. 77) für eine Auflistung aller Korrelationen zwischen dem DLC der Pferdewette und den Sicherheits-Ratings der Probanden. Mit steigendem DLC in der Pferdewette geben sich die Probanden mit einer geringeren Sicherheit zufrieden, dass (Szenario 5) eine Therapie gegen Heuschnupfen auch hilft (Pearson's $r = -.45$) und dass (Szenario 8) ein Traumpartner in einer Single-Börse auch wirklich zu einem Treffen erscheint (Pearson's $r = -.28$).

Es ist möglich, dass die Sicherheits-Ratings für die Entscheidungssituationen nur bedeutsam sind, wenn die jeweilige Entscheidungssituation für die Probanden im realen Leben auch relevant ist. Es wurden daher 13 lineare Modelle mit dem DLC der Pferdewette als Kriterium und dem Sicherheits-Rating und dem Rating der Relevanz der Situation für die Probanden im realen Leben als Prädiktoren gerechnet; die lineare Modelle sind hier nicht aufgeführt. Die linearen Modelle haben *nicht* die Vorzeichen der Beta Gewichte der Sicherheit-Ratings zu den Vorzeichen der Korrelationen verändert und das lineare Modell wurde „nur“ für das Szenario 5 und 8 signifikant, so dass hier nur die Korrelationen berichtet werden.

Die über die Aufgaben und Probanden gemittelte Punktzahl im Bilderrätsel liegt bei 76.5 bei einer Streuung über die Probanden von 4.8 und einer Spannweite über die Probanden von 65.3 bis 90.5.

Die Hypothese über den negativen Zusammenhang zwischen dem DLC der Pferdewette und der Punktzahl der Bilderrätsel-Aufgabe kann nicht bestätigt werden. Es zeigt sich zwar deskriptiv, dass die Probanden mit einem hohen DLC dazu tendieren, mehr Kästchen in der Bilderrätsel-Aufgabe aufzudecken, bevor sie einen Tipp über das Objekt unter der Maske abgeben, der Zusammenhang ist jedoch nicht signifikant von 0 verschieden. Cronbach's α der Punktezahlen der 40 Durchgänge des Bilderrätsels liegt bei .89. Die hohe interne Konsistenz der Aufgabe spricht dafür, dass die 40 Durchgänge das gleiche Konstrukt erfassen, so dass die Rechnung mit den über die Durchgänge für jeden Probanden gemittelten Punktzahlen gerechtfertigt ist. Die Korrelation zwischen dem DLC der Pferdewette und der mittleren Punktzahl des Bilderrätsels liegt bei Pearson's $r = -.19$, $t(49) = -1.39$, $p = .08$.

Zusammenfassend zeigt sich in dieser Studie, dass das DLC der Mess-

Table 2.2. Mittelwerte, Streuungen und Korrelationen (Pearson's r) der Sicherheits-Ratings der 13 Entscheidungssituationen mit dem DLC der Pferdewette.

| Entscheidungssituation | Sicherheits-Rating | | |
|----------------------------|--------------------|------|---------------------------|
| | M | SD | Pearson's r mit DLC Pferd |
| 1 Wetter Ausflug | 63.1 | 22.1 | -.06 |
| 2 Wechsel Studienort | 75.4 | 27.3 | -.03 |
| 3 Investition Immobilien | 80.8 | 28.9 | -.02 |
| 4 Zuverlässigkeit Ebay | 81.7 | 28.2 | -.05 |
| 5 Wirkung Therapie | 72.6 | 21.6 | -.45* |
| 6 Wirkung Impfung | 68.9 | 29.0 | -.11 |
| 7 Wechsel Job | 78.7 | 24.2 | -.03 |
| 8 Treffen Traumpartner | 58.6 | 29.7 | -.28* |
| 9 Organisation Party | 72.7 | 22.3 | -.08 |
| 10 Wohnung mit Partner | 62.8 | 25.2 | -.16 |
| 11 Unruhen Reiseland | 75.0 | 28.1 | .24 |
| 12 Energiesparen Hausbau | 80.4 | 24.1 | -.16 |
| 13 Funktionalität Prothese | 74.8 | 22.4 | -.06 |

Anmerkungen: *p < .05.

phase mit dem DLC der Testphase hoch korreliert ist, Probanden mit einem hohen DLC in der Pferdewette auch ein hohes DLC im Kriminalfall haben, *geringere* Sicherheitsratings in den Entscheidungssituation abgeben und vor Tippabgabe - zumindest deskriptiv - mehr offene Kästchen in der Bilderrätselaufgabe vor einer Entscheidung aufdecken.

2.1.4. Diskussion

Die prädiktive Kraft des DLC Modells sowie die hohe Korrelation zwischen dem DLC der Messphase und Testphase zeigt, dass die Messaufgabe das DLC reliabel erfasst. Die weiterhin hohe Korrelation zwischen dem DLC der Pferdewette und des Kriminalfalls zeigt, dass das DLC unabhängig von den Merkmalen der Oberfläche und spezifischen Parameter-Ausprägungen ist, was dafür spricht, dass ein spezifisches DLC nicht ein Artefakt der Einkleidung der Mess-Aufgabe ist.

Die höhere Anzahl der als DLC Nutzer klassifizierten Probanden im

2. Reliabilität und Validität

Kriminalfall (88%) versus Pferdewette (66%) kann verschiedene Gründe haben: 1) Es ist möglich, dass die Probanden ihre Stoppregel über die beiden Aufgaben hinweg verändert haben; 2) es ist jedoch auch möglich, dass die Probanden ihre Stoppregel über die beiden Aufgaben hinweg *nicht* verändert haben und die Kriminalaufgabe das „bessere“ Messinstrument ist, d. h., weniger fehlerbehaftet ist und mehr Konstruktvarianz misst.

Eine Antwort dazu kann in dieser Studie nicht gegeben werden. Es ist dafür notwendig, die Bearbeitungs-Reihenfolge der DLC Pferdewette und der DLC Kriminalaufgabe experimentell zu variieren, um Veränderungen, die auf eine Änderung der Strategieranwendung hindeuten, ausschließen zu können.⁶

Für eine Änderung der Strategiewahl spricht das Muster der Strategieklassifikation in den beiden DLC Aufgaben; siehe Tabelle 2.1 (S. 75). Alle vier Anwender einer MRDM Strategie wechseln zu einer weniger aufwendigen⁷ DLC Strategie, vier Probanden wechseln von einer uneindeutigen Strategie zur DLC Strategie, was möglicherweise auf die Ausbildung einer konsistenten Anwendung einer Strategie über die Studie hinweg zurückzuführen ist, und zwei Probanden wechseln von einer DLC Strategie zu einer Zufallsstrategie, was möglicherweise auf eine sinkende Motivation über die Studie hinweg zurückzuführen ist.

Die Korrelationen mit den Sicherheitseinschätzungen in den 13 Entscheidungssituationen sind - bis auf zwei Szenarien - nicht signifikant. Die Richtung der beiden signifikanten Korrelationen fällt erwartungs-*inkonsistent* aus.

Es stellt sich die Frage, inwieweit die Szenarien ein Konstrukt abbilden, das konzeptuell ähnlich zu dem DLC der Pferdewette ist. Es handelt sich bei den Szenarien um einen anderen Entscheidungstyp. Probanden fällen Präferenz-Entscheidungen: Die Konsequenzen der Entscheidungs-Option sind bekannt und die individuell notwendige Höhe der Wahrscheinlichkeit des Eintretens der positiven Konsequenzen für die Wahl der Option wird abgefragt (Glöckner, 2006, S. 17). Es ist fraglich, ob das DLC, das das Anspruchsniveau an gewünschter Urteilsicherheit abbilden soll, mit dem Sicherheitsurteil in das Auftreten einer positiven Konsequenz zusammenhängt. So muss die Abgabe eines hohen Sicherheitsurteils nicht zwangsläufig ein hohes individuelles Sicher-

⁶Die Reihenfolge wurde aus ökonomischen Gründen und inhaltlichen Gründen - es soll ja von dem DLC der Pferdewette als Prädiktor auf das DLC einer weiteren Aufgabe geschlossen werden - nicht variiert.

⁷weniger aufwendig = geringere Cue Konsultation und damit geringere Suchkosten

heitsbedürfnis bedeuten, sondern kann auch „nur“ aus einem geringen subjektiven Wert⁸ der Konsequenzen folgen.

Diese Argumentation spricht jedoch für Nullkorrelationen zwischen dem DLC und den Szenarien. Es ist dagegen zu beachten, dass mit steigender Anzahl der statistischen Tests am gleichen Datensatz die Wahrscheinlichkeit einer inkorrekten Entscheidung durch die Akkumulation des α -Fehlers ansteigt. Bei 13 Aufgaben führt das bei einer Irrtumswahrscheinlichkeit von 5% zu .65 Fehlklassifikationen also in etwa einem Szenario, so dass zwei signifikante Korrelation bei 13 Korrelationen mit Vorsicht interpretiert werden müssen und die Stabilität der Befunde bezweifelt werden muss. Dagegen lässt sich jedoch halten, dass die Test-Power mit .71 bei einem erwarteten mittleren Effekt ($r = .5$) gering ausfällt, die Wahrscheinlichkeit für einen β -Fehler also erhöht ist.

Das DLC korreliert nicht signifikant mit der Punktzahl im Bilderätsel, was jedoch nicht unbedingt auf ein konzeptuelles sondern ein statistisches Problem der Testpower zurückzuführen ist.⁹ Ein positiver Befund hätte es erlaubt, die Bilderrätsel Aufgabe zur Erfassung der Reliabilität (Interkorrelation) oder zur Stärkung der Reliabilität (durch Verrechnung der Werte) des DLC ohne großen zeitlichen Aufwand und ohne die mit einer Zweiterhebung der Pferdewette verknüpften methodischen Probleme zu verwenden.

Zusammenfassend kann festgehalten werden, dass die Aufgabe der Pferdewette das DLC hinreichend reliabel erfasst. Es ist nun nach Klärung dieser Grundvoraussetzung überhaupt erst möglich, die Validität des DLC in den folgenden Studien zu untersuchen.

2.2. Validität des DLC

In der Studie soll untersucht werden, ob das DLC zeitlich stabil ist und in welcher Relation das DLC zum Forschungsprogramm der adaptiven Werkzeugkiste (Gigerenzer et al., 1999) steht. Es wird somit die grundlegende Annahme zum DLC getestet, dass es sich beim DLC um eine Persönlichkeitseigenschaft handelt, die mit verwandten Konstrukten aus der Entscheidungspsychologie auch tatsächlich zusammenhängt.

⁸siehe *subjective utility theory* (Jungermann, Pfister & Fischer, 2010)

⁹Die Korrelation zwischen dem DLC des Kriminalfalls und der Punktezahl im Bilderrätsel fällt mit $r = -.26$, $t(49) = -1.89$, $p = .03$ signifikant aus, so dass die Bilderräselaufgabe in einer folgenden Studie noch einmal untersucht wird.

2. Reliabilität und Validität

Dazu wird neben dem DLC in der Pferdewette auch das Entscheidungsverhalten der Probanden im Börsenspiel, das im Theorieteil vorgestellt wurde, erhoben. In einem Vergleich der beiden Aufgaben ist es dann möglich, klassische Prozessmaße (Payne et al., 1988) und die Strategieklassifikation (Bröder & Schiffer, 2003a) mit dem DLC in Verbindung zu bringen. Es ist dann möglich, das DLC im Konstrukttraum klassischer Entscheidungsmodelle und -prozesse zu verorten.

2.2.1. Methoden

Die Studie bestand aus fünf Phasen. In der ersten beiden Phase wurde das DLC gemessen und das Entscheidungsverhalten der Probanden im Börsenspiel erhoben. In den folgenden beiden Phasen bearbeiteten die Probanden eine Woche später wiederholt die gleichen Aufgaben (DLC und Börsenspiel). In der abschließenden Phase wurden Persönlichkeitsvariablen erhoben. Im Folgenden werden die Aufgaben der einzelnen Phasen genauer beschrieben.

Die Probanden wurden schriftlich über ihre Rechte als Teilnehmer an einer psychologischen Studie aufgeklärt. Die Durchführung aller Aufgaben fand am Computer statt. Die Probanden erhielten für ihre Teilnahme für ihr Studium notwendige Versuchspersonenstunden gutgeschrieben und die drei Probanden mit der höchsten Punktzahl (s. u.) zusätzlich einen Kinogutschein.

Die Probanden bearbeiteten wiederholt in zwei Sitzungen im Abstand von 7 +/- 1 Tag die Pferdewette und das Börsenspiel.

Die Probanden bearbeiteten zuerst die 71 Durchgänge der Pferdewette, die in Kapitel 1.3.2.1 (S. 60) und Anhang A.2 (S. 171) beschrieben ist.

Dann bearbeiteten die Probanden 80 Durchgänge des Börsenspiels, das auf S. 38 in Kapitel 1.2.5 beschrieben wird.¹⁰

Die Felder der 3 × 4 Optionen-Cues Matrix des Börsenspiels sind am Anfang eines Durchgangs verdeckt; siehe Abbildung 1.2 (S. 39) und Abbildung 1.3 (S. 40) für die Oberfläche des Börsenspiels. Die Probanden können, indem sie mit dem Mauszeiger ein Feld anklicken, die Information eines Hinweisreizes abrufen. Die Information ist so lange sichtbar, wie die Probanden mit dem Mauszeiger im jeweiligen Informations-Feld bleiben.

Der prozentuale Gewinn bzw. Verlust jeder Aktie wurde über eine lineare Verknüpfung der Cues berechnet (Platzer, 2008, S. 77): Prozent

¹⁰siehe Anhang B.2 (S. 176) für die Instruktion des Börsenspiels

Aktie = $32 \times \text{cue}_1 + 21 \times \text{cue}_2 + 15 \times \text{cue}_3 + 11 \times \text{cue}_4$. „Cue“ kann dabei die Werte 1 für die Ausprägung „ja“ und -1 für die Ausprägung „nein“ annehmen. Damit die Umwelt nicht deterministisch ist, wurde auf den prozentualen Gewinn bzw. Verlust einer Aktie eine Zufallszahl, die einen Wert von $[-5, -4, \dots, 0, \dots, 4, 5]$ annehmen konnte, addiert.

Zwar ist die Umwelt kompensatorisch und begünstigt daher MRDM Strategien, jedoch wird dies für ORDM Strategien durch die Suchkosten von 2% pro aufgedecktem Cue unter dem gegebenen Cue-Muster der Durchgänge ausgeglichen. Die Umwelt begünstigt mit der gegebenen Gewinnstruktur also weder ORDM noch MRDM Strategien: Bei konsequenter Anwendung einer *Take The Best* Strategie versus *Franklin's Rule* Strategie erhalten die Probanden unter den beschriebenen Umweltbedingungen 12,939 Euro versus 12,934 Euro Spielgeld (Platzer, 2008, S. 77).

In der zweiten Sitzung wurde dann das *Bedürfnis nach kognitiver Geschlossenheit* (Schlink & Walther, 2007) und der *regulatorische Fokus* (Werth & Förster, 2007) am Computer erhoben. Die Beschreibung der Skalen, Hypothesen des Zusammenhangs mit dem DLC und die Ergebnisse, werden - wie schon im Theorieteil erörtert - im abschließenden Kapitel 4.2 (S. 135) dargestellt.

In der letzten Phase der Studie wurden demographische Daten der Probanden erhoben.

2.2.1.1. Stichprobe

Es nahmen 51 Probanden an der Studie teil. Davon waren 4 männlich und 38 weiblich, zu den übrigen 9 Probanden lagen aufgrund eines Erhebungsfehlers keine demographische Daten vor. Das Alter der Probanden lag zwischen 15 und 58 Jahren bei einem Mittelwert von 28,5 Jahren und einer Streuung von 11,6 Jahren.

2.2.1.2. Prozessmaße einer Entscheidung

Im Börsenspiel wurde über die Entscheidungsdurchgänge hinweg die folgenden Prozessmaße einer Entscheidung aufgezeichnet (Payne et al., 1988).

Es wurde die mittlere Anzahl der Cues, die die Probanden über die Durchgänge hinweg aufdecken, erhoben. Anwender einer ORDM Strategie decken tendenziell weniger Cues auf als Anwender einer MRDM Strategie: Die strikte Anwendung einer ORDM Strategie führt z. B., wenn der valideste Cue zwischen den Optionen diskriminiert, zu drei

2. Reliabilität und Validität

aufgedeckten Cue-Feldern: Das sind die Cue Felder des validesten Cues für jede Option. Die strikte Anwendung einer MRDM Strategie führt dazu, dass immer alle Cue-Felder inspiziert werden. Da eine strikte Anwendung einer Strategie empirisch unwahrscheinlich ist, wird folgend der Begriff „tendenziell“ verwendet.

Es wurde die mittlere Inspektionszeit, die zwischen dem Anklicken eines Cue-Feldes und dem Verlassen des Cue-Feldes liegt, gemessen. Anwender einer ORDM Strategie inspizieren die Cues tendenziell länger als Anwender einer MRDM Strategie.¹¹

Es wurde die mittlere prozentuale Zeit, die über die Durchgänge hinweg mit der Inspektion des validesten Cues verbracht wurde, gemessen. Anwender einer ORDM Strategie verwenden tendenziell prozentual mehr Zeit für die Inspektion des validesten Cue als Anwender einer MRDM Strategie. So verbringt ein strikter Anwender einer ORDM Strategie, wenn der valideste Cue über alle Aufgaben hinweg diskriminiert, 100% der Inspektionszeit auf den validesten Cue.

Es wurde erhoben, wie oft alternativ- versus attributweise die Optionen-Cues Matrix abgesucht wurde. Ein Anwender einer MRDM Strategie verrechnet tendenziell die gesamte Information einer Option und zeigt daher ein Suchverhalten über die Cue-Ausprägungen einer Alternative hinweg (= alternativweise Suche). Ein Anwender einer ORDM Strategie tendiert dazu, die Cues nach ihrer Validität abzusuchen, um einen Cue zu finden, der zwischen den Optionen diskriminiert (= attributweise Suche). Ein Anwender einer ORDM Strategie zeigt daher tendenziell mehr Suchverhalten über die Cue-Ausprägungen von Optionen. Kritisch daran ist, dass nicht klar ist, wie verschiedene - möglicherweise sich widersprechende - Prozess- und Outcome-Maße einer Entscheidung zu einem Index verrechnet werden können; siehe dazu auch Bröder (2000a). Das bedeutet für diese Studie, dass die Korrelationen mit dem DLC durch eine niedrige Konstruktvalidität der Prozessmaße und eine niedrigen Stabilität der Prozessmaße niedrig ausfallen kann. Daher wurden die Interkorrelationen zwischen den Prozessmaßen und

¹¹Es ist nicht direkt einleuchtend, warum Anwender einer MRDM Strategie mehr Zeit für die Inspektion eines Cues verwenden sollten. Payne et al. (1988, S. 543, Hervorhebung nicht im Original) schreiben dazu: „A second measure, which is related to the amount of *processing effort* [...], is the average time spent per item of information acquired“. Siehe dazu auch den Überblick von Horstmann et al. (2009, S. 337) zum positiven Zusammenhang zwischen der Zeit der Augenfixation einer Information und der Verarbeitungstiefe einer Information und der Hypothese, dass MRDM Strategien durch die Verrechnung von Cues Entscheidungen eine höhere Verarbeitungstiefe abverlangen.

die zeitliche Stabilität der Prozessmaße bestimmt.

Der Grad der alternativ- versus attributweisen Suche wurde im Strategie-Index nach Payne et al. (1988, S. 543) verrechnet:

Strategie-Index = $\frac{H(\text{Alternativ}) - H(\text{Attribut})}{H(\text{Alternativ}) + H(\text{Attribut})}$. Der Strategie-Index kann einen Wert zwischen -1 und 1 annehmen. Ein negativer Wert deutet auf attributweise Suche hin, ein positiver Wert auf alternativweise Suche und ein Wert von 0 auf gleich viele attribut- wie alternative Suchwechsel in der Optionen-Cues Matrix.

2.2.1.3. Bayesianische Strategieklassifikation

Im Folgenden wird die konzeptuelle Idee der Bayesianischen Strategieklassifikation skizziert. In Anhang A.1 (S. 167) ist eine ausführlichere Beschreibung der Methode mit den mathematischen Details zu finden.

In der Bayesianischen Strategieklassifikation dieser Studie wird für die Strategien *Take The Best*, *Dawes Rule* und *Franklin's Rule* die erwarteten Wahlen über das Set von Entscheidungsaufgaben im Börsenspiel bestimmt. Dann wird für jede Strategie und jeden Probanden der Vektor mit den erwarteten Wahlen mit dem Vektor der tatsächlich beobachteten Wahlen verglichen. Es lässt sich dann unter bestimmten Annahmen (siehe Anhang A.1) in einem Passungsindex berechnen, welche Strategie die Probanden am wahrscheinlichsten verwendet haben. Darüber hinaus kann der Fehler, mit der eine Strategie angewendet wird, berechnet werden, so dass man nicht nur einen Index für die Passung sondern auch für die Konsistenz der Anwendung einer Strategie erhält.

2.2.2. Hypothesen

Das DLC ist ein zeitlich stabiles Konstrukt. Das DLC des ersten Messzeitpunktes ist mit dem DLC des zweiten Messzeitpunktes positiv korreliert.

Das DLC bildet ORDM und MRDM Suchverhalten einer Entscheidungs-Strategie ab. Das DLC ist positiv mit der mittleren Anzahl der aufgedeckten Cues korreliert, positiv mit der mittleren Inspektionszeit der Cues, negativ mit der mittleren prozentual auf den ersten Cue verwendeten Inspektionszeit und positiv mit dem Strategie-Index.

Das DLC bildet nicht nur das Suchverhalten sondern auch die Integrationsregel einer Entscheidungs-Strategie (nach der Bayesianischen

2. Reliabilität und Validität

Klassifikation) ab¹²: Das mittlere DLC fällt in der Gruppe der Probanden, die im Börsenspiel als *Take The Best* Nutzer klassifiziert wurden, geringer aus als das mittlere DLC der Probanden, die als Nutzer der *Dawes Rule* oder *Franklin's Rule* klassifiziert wurden.

2.2.3. Ergebnisse

Das mittlere DLC der Pferdewette liegt zu Zeitpunkt 1 bei .824 bei einer Streuung von .110. Das mittlere DLC der Pferdewette liegt zu Zeitpunkt 2 bei .788 bei einer Streuung von .103. Die Modelltestung gegen den Zufall¹³ ergibt, dass zum *ersten* Messzeitpunkt 34 von 50 (68%) Probanden als Anwender des DLC als Stopppregel der Informationssuche in den 20 Testdurchgängen der Pferdewette klassifiziert werden können. 2 Probanden können als Anwender einer ORDM-, 6 Probanden als Anwender einer MRDM- und 1 Probanden als Anwender einer Zufalls-Stopppregel klassifiziert werden. 7 Probanden lassen sich nicht eindeutig nur einem der Modelle zuordnen.

Die Modelltestung gegen den Zufall ergibt, dass zum *zweiten* Messzeitpunkt 40 von 50 (80%) Probanden als Anwender des DLC als Stopppregel der Informationssuche in den 20 Testdurchgängen der Pferdewette klassifiziert werden können. 0 Probanden können als Anwender einer ORDM-, 5 Probanden als Anwender einer MRDM- und 0 Probanden als Anwender einer Zufalls-Stopppregel klassifiziert werden. 5 Probanden lassen sich nicht eindeutig nur einem der Modelle zuordnen.

In Tabelle 2.3 (S. 85) sind die Häufigkeiten der Modellklassifikation für das DLC der Pferdewette zu den beiden Messzeitpunkten gegeneinander in einer Kreuztabelle abgetragen.

Die Hypothese über die signifikant positive Korrelation zwischen dem DLC zu Messzeitpunkt 1 und Messzeitpunkt 2 kann bestätigt werden¹⁴. Die Probanden, die ein höheres DLC zum ersten Zeitpunkt haben, haben auch ein signifikant höheres DLC zum zweiten Zeitpunkt. Die Korrelation liegt bei Pearson's $r = .61$, $t(49) = 5.33$, $p < .05$.

Die Hypothese über die signifikant positive Korrelation zwischen dem DLC und der mittleren Anzahl der aufgedeckten Cues im Börsenspiel kann für beide Messzeitpunkte bestätigt werden. Probanden mit einem

¹²Das ist eine gewagte Hypothese, da das Aufdecken mehrerer Cues - ein MRDM Suchverhalten - nicht zwangsläufig mit der Integration aller Cues zu einer Entscheidung einhergehen muss (Bröder, 2000a).

¹³siehe S. 64 in Kapitel 1.3.2.2 für Details zur Testung gegen den Zufall

¹⁴Die Power liegt bei einer erwarteten mittleren Effektstärke von Pearson's $|r| = .3$ und einer Stichprobengröße von 51 und einem $\alpha = .05$ bei $1 - \beta = .72$.

Tabelle 2.3. Kreuztabelle der Modellklassifikation über die 20 Testdurchgänge des DLC zu Zeitpunkt t 1 und Zeitpunkt t 2.

| Modell- klassifikation t 1 | Modellklassifikation t 2 | | | | | Σ |
|----------------------------------|--------------------------|------|------|--------|-------------|----------|
| | DLC | ORDM | MRDM | Zufall | Uneindeutig | |
| DLC | 30 | 0 | 2 | 0 | 2 | 34 |
| ORDM | 2 | 0 | 0 | 0 | 0 | 2 |
| MRDM | 2 | 0 | 3 | 0 | 1 | 6 |
| Zufall | 1 | 0 | 0 | 0 | 0 | 1 |
| Uneindeutig | 5 | 0 | 0 | 0 | 2 | 7 |
| Σ | 40 | 0 | 5 | 0 | 5 | (50) |

Anmerkungen: n = 50, da bei einem Probanden die Daten zur Testphase des DLC aufgrund eines Computerfehlers nicht aufgezeichnet wurden.

hohen DLC, decken auch signifikant mehr Cues im Börsenspiel auf. Die Korrelation liegt zu Messzeitpunkt 1 bei Pearson's $r = .34$, $t(49) = 2.54$, $p < .05$ und zu Messzeitpunkt 2 bei Pearson's $r = .42$, $t(49) = 3.29$, $p < .05$.

Die Hypothese über die signifikant positive Korrelation zwischen dem DLC und der mittleren Inspektionszeit der Cues im Börsenspiel kann nicht bestätigt werden. Die Probanden, die ein hohes DLC haben, haben keine signifikant längere mittlere Inspektionszeit für die Cues im Börsenspiel. Die Korrelation zwischen dem DLC und der mittleren Inspektionszeit liegt zu Messzeitpunkt 1 bei Pearson's $r = .12$, $t(49) = 0.85$, $p = .19$ und zu Messzeitpunkt 2 bei Pearson's $r = -.15$, $t(49) = -1.07$, $p = .85$.

Die Hypothese über die signifikant negative Korrelation zwischen dem DLC und der prozentual auf den ersten Cue verwendeten Inspektionszeit im Börsenspiel kann nicht bestätigt werden. Die Probanden, die ein hohes DLC haben, verwenden nicht signifikant weniger prozentuale Inspektionszeit auf den validesten Cue im Börsenspiel. Die Korrelation zwischen dem DLC und der prozentual auf den ersten Cue verwendeten Inspektionszeit liegt zu Messzeitpunkt 1 bei Pearson's $r = -.01$, $t(49)$

2. Reliabilität und Validität

= -0.03, $p = .48$ und zu Messzeitpunkt 2 bei Pearson's $r = -.22$, $t(49) = -1.6$, $p = .057$.

Die Hypothese über die signifikant positive Korrelation zwischen dem DLC und dem Strategie-Index im Börsenspiel kann nicht bestätigt werden. Die Probanden, die ein hohes DLC haben, tendieren nicht dazu, mehr alternativ- als attributweise die Optionen-Cues Matrix abzusuchen. Der Zusammenhang zeigt deskriptiv sogar in die hypotheseninkonforme Richtung. Die Korrelation zwischen dem DLC und dem Strategie-Index liegt zu Messzeitpunkt 1 bei Pearson's $r = -.27$, $t(49) = -2.00$, $p = .97$ und zu Messzeitpunkt 2 bei Pearson's $r = -.13$, $t(49) = -0.94$, $p = .82$.

In Tabelle 2.4 (S. 87) sind neben den berichteten hypothesenrelevanten Korrelationen auch alle übrigen Interkorrelationen zwischen den Prozessmaßen für die beiden Messzeitpunkte und die Korrelationen der Prozessmaße zu Messzeitpunkt 1 und Messzeitpunkt 2 abgetragen. Auffällig ist hier z. B. die signifikant negative Korrelation mit einem Pearson's r von $-.53$ zwischen der mittleren Anzahl der aufgedeckten Cues (PT 1) und dem Strategie-Index (PT 4) zu Messzeitpunkt 1 (unterhalb der Diagonalen), d. h., dass Probanden, die viele Cues aufdecken - also ein MRDM Suchverhalten zeigen -, *unerwarteterweise* dazu tendieren, über die Attribute hinweg in der Optionen-Cues Matrix zu suchen - ein Verhalten das mit ORDM Strategien assoziiert ist. Eine signifikante Korrelation konnte jedoch zu Messzeitpunkt 2 nicht repliziert werden (siehe oberhalb der Diagonalen).

Tabelle 2.4. Mittelwert, Standardabweichung, t-Test auf Mittelwertsdifferenzen von Messzeitpunkt 1 zu Messzeitpunkt 2, Effektstärke δ'' der Mittelwertsdifferenzen¹⁵ für das DLC und die vier Prozessmaße einer Entscheidung im Börsenspiel, die Korrelation aller Variablen über die beiden Messzeitpunkte (in der Diagonalen in Klammern) und die Interkorrelation der Variablen zu Messzeitpunkt 1 (unter der Diagonalen) und zu Messzeitpunkt 2 (über der Diagonalen).

| Maß | t 1 | | t 2 | | t(50) | δ'' | Pearson's r | | | | |
|------|------|------|------|------|--------|------------|-------------|--------|--------|--------|--------|
| | M | SD | M | SD | | | DLC | PT 1 | PT 2 | PT 3 | PT 4 |
| DLC | .824 | .110 | .788 | .103 | 2.75* | .386 | (.61*) | .42* | -.15 | -.22 | -.13 |
| PT 1 | 5.4 | 1.8 | 5.8 | 1.7 | -2.86* | -.401 | .34* | (.85*) | -.20 | -.63* | -.24 |
| PT 2 | 343 | 157 | 260 | 118 | 5.93* | .830 | .12 | -.09 | (.77*) | .12 | .10 |
| PT 3 | .41 | .15 | .42 | .15 | -0.34 | -.048 | -.01 | -.38* | -.07 | (.72*) | -.22 |
| PT 4 | -.54 | .46 | -.58 | .41 | 0.87 | .122 | -.27 | -.53* | .04 | -.04 | (.70*) |

Anmerkungen: PT 1 = Mittlere Anzahl aufgedeckter Cues, PT 2 = Mittlere Inspektionszeit, PT 3 = Prozentualer Anteil der auf den validesten Cue verwendeten Zeit, PT 4 = Strategie-Index, t 1 = Messzeitpunkt 1, t 2 = Messzeitpunkt 2, * $p < .05$, alle Korrelationen - bis auf die in den Hypothesen genannten - wurden zweiseitig auf Signifikanz getestet.

2. Reliabilität und Validität

Es zeigt sich, dass die Prozessmaße einer Entscheidung zeitstabil sind; das mittlere Pearson's r liegt über die Maße hinweg bei .76 bei einer Streuung von .06.

Um die Interkorrelation der Prozessmaße zu kontrollieren und die Prädiktionskraft der Maße für das DLC zu ermitteln, wurde *ein* lineares Modell mit den über die beiden Messzeitpunkte gemittelten Prozessmaßen als Prädiktoren und dem über die beiden Messzeitpunkte gemittelten DLC als Kriterium gerechnet.

Tabelle 2.5. Multiple lineare Regression mit dem über die beiden Messzeitpunkte gemittelten DLC als Kriterium und den über die beiden Messzeitpunkte gemittelten vier Prozessmaßen einer Entscheidung als Prädiktoren.

| Prädiktor | B | SEB | β | t | p |
|-----------|--------|--------|---------|-------|-------|
| Intercept | .560 | .113 | - | 4.92 | < .05 |
| PT 1 | .027 | .012 | .42 | 2.21 | < .05 |
| PT 2 | < .001 | < .001 | .08 | 0.60 | .54 |
| PT 3 | .137 | .135 | .17 | 1.00 | .31 |
| PT 4 | -.032 | .045 | -.11 | -0.71 | .47 |

Anmerkungen: $R^2 = .19$, $F(4,46) = 2.69$, $p < .05$; PT 1 = Mittlere Anzahl aufgedeckter Cues, PT 2 = Mittlere Inspektionszeit, PT 3 = Prozentualer Anteil der auf den validesten Cue verwendeten Zeit, PT 4 = Strategie-Index.

In Tabelle 2.5 (S. 88) sind die Ergebnisse zur Regression dargestellt. Die mittlere Anzahl der aufgedeckten Cues im Börsenspiel ist mit einem β von .42 ($t(49) = 4.92$, $p < .05$) der einzige signifikante Prädiktor für das DLC. Die Prädiktoren können 19% der Varianz im DLC aufklären.

Die Summe der geschätzten Fehlerwahrscheinlichkeiten in der Strategieanwendung der Probanden, das ist $\hat{\epsilon}_k + \hat{\phi}_k$, liegt im Mittelwert bei .50 mit einer Streuung von .09 zu Zeitpunkt 1 und einem Mittelwert bei .49 mit einer Streuung von .08 zu Zeitpunkt 2. Das bedeutet, dass die Probanden im Mittel die Strategie, die sie im Vergleich zu den getesteten Strategien (siehe Tabelle A.1, S 168) am wahrscheinlichsten angewendet haben, in $\approx 50\%$ der Durchgänge *inkonsistent* anwenden. Es ist fraglich, ob dieser geringe Anteil strategiekonsistenter Wahlen

¹⁵Notation nach Eid, Gollwitzer und Schmitt (2010, F 12.8, S. 353): $\delta'' = \frac{\mu_D}{\sigma_D}$ (vgl. mit Cohen, 1988)

für die Anwendung der Strategien spricht oder ob es nicht wahrscheinlicher ist, dass die Probanden, die einen hohen Fehlerwert haben, (eine) Strategie/n verwenden, die *nicht* zu dem Set der getesteten Strategien gehört/en oder eine Raten-“Strategie“ verfolgten. Die folgenden Ergebnisse der Strategieklassifikation basieren daher nur auf den Probanden, die einen Fehlerwert $< .5$ haben. Das Kriterium erfüllen 21 Probanden zu beiden Messzeitpunkten und 32 Probanden zu einem Messzeitpunkt.

Die Hypothese, dass das DLC auch die Integrationsregel einer Strategie abbildet, kann nicht bestätigt werden. Es ist zu beiden Messzeitpunkten nicht der Fall, dass Probanden, die als Nutzer einer ORDM Strategie - das ist *Take The Best* - versus MRDM Strategie - das sind *Dawes Rule* und alle drei Arten von *Franklin's Rule* - klassifiziert werden können, ein signifikant niedrigeres DLC haben; siehe Tabelle 2.6 (S. 89). Die Mittelwertsdifferenzen zeigen zwar deskriptiv in die richtige Richtung, sie sind zu Messzeitpunkt 1 mit einem $t(25) = -0.46$, $p = .32$ und zu Messzeitpunkt 2 mit einem $t(25) = -0.38$, $p = .35$ jedoch nicht signifikant.¹⁶

Tabelle 2.6. Mittelwert des DLC, Streuung des DLC und Anzahl der Probanden, die im Börsenspiel als *Take The Best* (= ORDM) versus *Franklin's Rule* und *Dawes Rule* (= MRDM) Strategienutzer identifiziert wurden; t-Test und Cohen's d zum Mittelwertsvergleich des DLC zwischen den beiden Gruppen für den jeweiligen Messzeitpunkt t1 und t2.

| Messung | ORDM | | | MRDM | | | t | df | p | Cohen's d |
|---------|------|------|----|------|------|----|-------|----|-----|-----------|
| | M | SD | n | M | SD | n | | | | |
| t1 | .830 | .111 | 17 | .849 | .072 | 10 | -0.46 | 25 | .32 | 0.19 |
| t2 | .806 | .092 | 19 | .821 | .076 | 8 | -0.38 | 25 | .35 | 0.17 |

Anmerkungen: Es wurde nur mit den Daten der Probanden gerechnet, die zum jeweiligen Messzeitpunkt einen Fehlerwert (= strategieinkonsistente Optionswahlen) $< .5$ hatten.

Als Zusatzanalyse sind in Tabelle 2.7 (S. 90) die Klassifikation zu

¹⁶Die Power liegt bei einer erwarteten mittleren Effektstärke von Cohen's d = .5 und einer Stichprobengröße von 17 für ORDM resp. 10 für MRDM zu Messzeitpunkt 1 und einer Stichprobengröße von 19 für ORDM resp. 8 für MRDM zu Messzeitpunkt 2 und einem $\alpha = .05$ jedoch auch *nur* bei $1 - \beta = .33$ zu Messzeitpunkt 1 und $1 - \beta = .31$ zu Messzeitpunkt 2.

2. Reliabilität und Validität

Messzeitpunkt 1 und Messzeitpunkt 2 gegeneinander abgetragen. Es wird nicht zwischen den verschiedenen Arten von *Franklin's Rule* unterschieden.

Tabelle 2.7. Kreuztabelle Häufigkeiten der Strategieklassifikation zu Messzeitpunkt 1 und Messzeitpunkt 2 für 21 Probanden, die zu beiden Messzeitpunkten einen Fehlerwert (= strategieinkonsistente Optionswahlen) < .5 haben.

| Strategie-Klassifikation t 1 | Strategie-Klassifikation t 2 | | | Σ |
|------------------------------|------------------------------|-----------------|------------|----------|
| | Take The Best | Franklin's Rule | Dawes Rule | |
| Take The Best | 10 | 1 | 0 | 11 |
| Franklin's Rule | 4 | 2 | 0 | 6 |
| Dawes Rule | 1 | 1 | 2 | 4 |
| Σ | 15 | 4 | 2 | (21) |

Anmerkungen: t 1 = Messzeitpunkt 1, t 2 = Messzeitpunkt 2.

Es zeigt sich, dass von den 21 Probanden, die zu beiden Messzeitpunkten einen Fehlerwert < .5 haben, 14 Probanden (= 66%) bei der gleichen Strategie bleiben. Die Mehrheit der Probanden, die ihre Strategie wechseln - das sind 5 von insgesamt 7 Probanden - schwenken von einer MRDM Strategie zu Messzeitpunkt 1 zu einer weniger aufwändigen ORDM Strategie - das ist *Take The Best* - zu Messzeitpunkt 2 um.

Zusammenfassend zeigt sich in dieser Studie, dass das DLC zu beiden Zeitpunkten korreliert ist, mit der Anzahl der aufgedeckten Cues im Börsenspiel korreliert und nicht die Art der Integration der Cues abbilden kann.

2.2.4. Diskussion

Die (moderate) zeitliche Stabilität des DLC erfüllt eine zentrale Bedingung für ein Persönlichkeitskonstrukt. Es ist nun notwendig zu untersuchen, ob das DLC mit situationalen Manipulationen, die auf das zugrundeliegende angenommene Konstrukt des *gewünschten Anspruchsniveaus an Urteilsicherheit* (Hausmann-Thürig, 2004) einwirken, interagiert, und ob das DLC mit verwandten Persönlichkeitsmaßen kor-

reliert ist. Dieser nächste Schritt ist notwendig, um zu prüfen, ob die Bedeutung des DLC nicht nur auf die Pferdewette und das inhaltlich nahe Börsenspiel limitiert ist, sondern über den Kontext dieser Aufgabentypen hinaus psychologische Bedeutung - d. h., ein Maß für das gewünschte Anspruchsniveau an Urteilsicherheit zu sein - hat.

Die hohe zeitliche Stabilität der Prozessmaße und die für die Mehrheit (Fehlerwert $< .5$) der Probanden stabile Strategieklassifikation lässt darauf schließen, dass Entscheidungsprozesse und die Entscheidung selbst auf eine stabile Anwendung einer Strategie zurückgehen. Es konnten somit über die von Bröder und Schiffer (2006a) berichteten Routine-Effekte für die Strategieklassifikation hinaus Routine-Effekte für weitere Maße einer Entscheidung und für einen längeren Erhebungszeitabstand von einer Woche gezeigt werden.

Die Korrelation des DLC mit der mittleren Anzahl der aufgedeckten Cues im Börsenspiel und die Befunde, dass das DLC nicht mit den anderen Prozessmaßen einer Entscheidung korreliert ist und sich das mittlere DLC zwischen den Anwendern von ORDM und MRDM Strategien nicht signifikant unterscheidet, legen nahe, dass das DLC den Abbruch der Informationssuche - jedoch *nicht* eine strategiespezifische Integration der Informationen (Gigerenzer & Goldstein, 1996) - erfasst.

Kritisch ist hier jedoch die hohe Fehlerrate in der Anwendung der Strategien¹⁷, die daraus resultierende hohe Dropout-Rate bei dem gesetzten Kriterium von Fehlerwert $< .5$ und die damit einhergehende geringe Power der Tests.

Mit Blick auf die Prozess-Bausteine einer Entscheidung in der adaptiven Werkzeugkiste, das sind die Such-, Stopp- und Entscheidungsregel (Gigerenzer & Goldstein, 1996), bildet das DLC also nur die Stoppregel ab. Über die Suchregel kann keine Aussage gemacht werden, da die Reihenfolge der Cues in der Pferdewette vorgegeben ist. Über die Entscheidungsregel kann keine Aussage gemacht werden, da ein bestimmter DLC Wert nicht mit einem Strategietyp einhergeht. Das DLC ist also kein genereller Proxy für die Prozess- *und* Outcome-Maße der Strategien der adaptiven Werkzeugkiste und bildet selbst innerhalb der Outcome-Maße nur die Anzahl der konsultierten Informationen ab.

81% der Varianz im DLC können nicht durch die Prozessmaße der Entscheidung aufgeklärt werden, so dass das DLC über die Prozessmaße einer Entscheidung hinaus etwas „Anderes“ misst. Ob es sich dabei um systematische Varianz (Konstruktvarianz) oder um unsystematische

¹⁷siehe zum Einfluss der Höhe des Fehlers auf die Reliabilität der Strategieklassifikation auch die Simulationsstudie von Bröder und Schiffer (2003a)

2. Reliabilität und Validität

sche Varianz (Fehlervarianz) handelt, sollen die folgenden Studien zeigen, in denen das DLC mit (theoretisch) konstruktrelevanten Manipulationen situational verändert und mit konstruktrelevanten Persönlichkeitsvariablen korreliert werden soll.

Die geringen (bzw. nicht vorhandenen) Korrelationen zwischen den Maßen können aber nicht nur auf Konstruktebene erklärt werden. Ein weiterer möglicher (trivialerer) Grund ist, dass die Messung des DLC und der Prozessmaße nicht reliabel sind. Wenn die Maße an sich nicht reliabel sind, ist auch keine hohe Interkorrelation der Maße zu erwarten. Dem ist jedoch entgegenzuhalten, dass sich das DLC in Studie 1 als reliables Maß herausgestellt hat und dass die Korrelation des DLC und der Prozessmaße in Studie 2 über die beiden Messzeitpunkte sehr wahrscheinlich niedriger ausgefallen wären, wenn es sich bei den geringen (bzw. fehlenden) Interkorrelation der Maße "nur" um ein Messproblem gehandelt hätte.

3. Situative Einflussfaktoren

In den folgenden Studien wird untersucht, ob sich das DLC durch situationale Manipulationen verändern lässt. Dazu werden in zwei Studien zwei unterschiedliche Arten des Primings verwendet. In weiteren zwei Studien werden auf zwei Arten ein hohes vs. niedriges *Need for Cognitive Closure* induziert (Kruglanski & Webster, 1996; Kruglanski, 2004). Alle situativen Manipulationen sollen ORDM vs. MRDM begünstigen, d. h. also ein niedriges vs. hohes DLC induzieren.

3.1. Priming

In den folgenden beiden Studien soll untersucht werden, ob das DLC mit situationalen Manipulationen interagiert. Dazu soll auf zwei Arten des Primings - nämlich konzeptuelles Priming und Mindest Priming (Bargh & Chartrand, 2000) - ORDM bzw. MRDM herbeigeführt werden.

3.1.1. Konzeptuelles Priming: Sätze bilden

In einer Aufgabe zum konzeptuellen Priming werden internale mentale Repräsentationen über eine Darbietung von Stimuli aktiviert (Bargh & Chartrand, 2000). Die aktivierte Repräsentation wirkt auf eine nachfolgende Zielaufgabe. Die Aktivierung einer spezifischen Repräsentation ist also die unabhängige Variable und die gemessene Wirkung der Repräsentation auf die Zielaufgabe die abhängige Variable einer experimentellen Priming Studie. Eine aktivierte Repräsentation kann dabei nicht nur Auswirkungen auf die Wahrnehmung und Beurteilung von Objekten der Umwelt haben (Devine, 1989; Higgins, Rholes & Jones, 1977), sondern auch auf das Verhalten wirken (Aarts & Dijksterhuis, 2003; Bargh, Chen & Burrows, 1996); siehe dazu den *Perception-Behavior Link* (Bargh & Williams, 2006; Dijksterhuis & Bargh, 2001).¹

¹siehe z. B. Förster und Liberman (2007) für einen Überblick zu Priming Effekten

3. Situative Einflussfaktoren

Um sicherzustellen, dass nur die Aktivierung der Repräsentation einen Einfluss auf die Zielaufgabe hat (Lombardi, Higgins & Bargh, 1987; Stapel & Suls, 2007), muss die Priming Aufgabe so gestaltet werden, dass für die Probanden der Zusammenhang zwischen der Priming Aufgabe und der Zielaufgabe nicht offensichtlich ist.

Eine bewährte Methode des supraliminalen Primings - das ist die Darbietung von Stimuli oberhalb der Wahrnehmungsschwelle - ist die *Scrambled Sentences* Technik (Bargh et al., 1996; Epley & Gilovich, 1999; Fishbach, Dhar & Zhang, 2006; Macrae & Johnston, 1998; Smith & Trope, 2006; Srull & Wyer, 1979; Walther, Müller & Schott, 2001). In einer *Scrambled Sentences* Priming Aufgabe werden Probanden in mehreren Durchgängen Wörter in Zufallsreihenfolge präsentiert, aus denen die Probanden grammatikalisch korrekte Sätze bilden sollen (Bargh & Chartrand, 2000, S. 259). Unter den Wörtern befinden sich kritische Begriffe, die mit der zu aktivierenden Repräsentation assoziiert sind.

So sollten z. B. Probanden in einer Studie von Bargh et al. (1996, Experiment 2) Sätze aus Wörtern bilden, die mit dem Konzept *alt* assoziiert sind wie z. B. Florida, einsam, sentimental, grau und Falten (Bargh et al., 1996, S. 236). Die so in der Priming Aufgabe aktivierte mentale Repräsentation *alt* führte in der Zielaufgabe zu Verhalten, das mit *alt* assoziiert ist: Die Probanden der Priming Bedingung benötigten mehr Zeit, um von der Tür des Labors zum Fahrstuhl des Gebäudes zu gelangen als Probanden, bei denen die alterstereotypen Begriffe der *Scrambled Sentences* Aufgabe durch neutrale Begriffe ersetzt wurden.

In einer Studie von Platzer (2008) wurde der Einfluss eines konzeptuellen Primings von ORDM und MRDM auf die Prozessmaße einer Entscheidung (Payne et al., 1988) und die Strategieklassifikation (Bröder & Schiffer, 2003a) im Börsenspiel untersucht.² Dazu bearbeiteten die Probanden eine *Scrambled Sentences* Aufgabe, in der Begriffe vorkamen, die mit ORDM oder MRDM Entscheidungsverhalten assoziiert sind. Es zeigte sich kein Einfluss des Primings auf die Prozessmaße und die Strategieklassifikation von Entscheidungsverhalten.

In der vorliegenden Studie wird die von Platzer (2008) entwickelte Priming Methode verwendet, um den Einfluss eines Primings von ORDM vs. MRDM auf das DLC zu untersuchen. Die Befunde von Platzer (2008) sprechen gegen die Wirksamkeit des Primings auf das Entscheidungsverhalten. Dies kann auf methodische Probleme, d.h., die spezifischen ORDM vs. MRDM Begriffe der Priming Aufgabe haben nicht gewirkt, oder konzeptuelle Probleme, d. h., spezifische Entschei-

²siehe S. 38 in Kapitel 1.2.5 für eine Beschreibung des Börsenspiels

dingsstrategien können nicht geprimet werden, zurückgeführt werden. Jedoch ist das DLC konzeptuell nicht identisch mit den Prozessmaßen einer Entscheidung und der Strategieklassifikation (siehe vorherige Studien). Weiterhin sind die Skaleneigenschaften des DLC zur Strategieklassifikation unterschiedlich: Das DLC ist metrisch vs. die Strategieklassifikation ist kategorial. Die höhere Differenzierbarkeit zwischen den Probanden beim DLC kann sensitiver für das Priming sein, wenn die interindividuelle Varianz im DLC bedeutsam ist. Abschließend ist die inhaltliche Einkleidung der Messaufgabe der Pferdewette zum Börsenspiel unterschiedlich, so dass eine Replikation der verwendeten Priming Methode berechtigt ist.

3.1.1.1. Methoden

Die Studie bestand aus vier Phasen. In der ersten Phase bearbeiteten die Probanden die Primingaufgabe. In der zweiten Phase wurde das DLC gemessen. Dann bearbeiteten die Probanden eine Manipulation-Check Aufgabe, die die Wirkung des Primings überprüfen sollte. In der letzten Phase wurden Persönlichkeitsvariablen erhoben. Im Folgenden werden die Aufgaben der einzelnen Phasen genauer beschrieben.

Die Probanden wurden schriftlich über ihre Rechte als Teilnehmer an einer psychologischen Studie aufgeklärt. Die Durchführung aller Aufgaben fand am Computer statt. Die Probanden erhielten für ihre Teilnahme für ihr Studium notwendige Versuchspersonenstunden gutgeschrieben und die drei Probanden mit der höchsten Punktzahl (s. u.) erhielten zusätzlich einen Kinogutschein.

Das konzeptuelle Priming von ORDM und MRDM wurde über die *Scrambled Sentences* Technik - die in der Instruktion zur Aufgabe als Testung der Sprachfertigkeit getarnt wurde - induziert (Bargh & Chartrand, 2000). In allen Bedingungen sollten die Probanden in 20 Durchgängen grammatikalisch korrekte Sätze aus vier von fünf dargebotenen Wörtern bilden. Zum Beispiel konnte aus den Wörtern „interessante, berücksichtigt, wurden, Indizien, gesammelt“ eines Durchgangs der neutralen Bedingung der syntaktisch korrekte Satz „Interessante Indizien wurden berücksichtigt“ gebildet werden. 12 von den 20 Durchgängen beinhalteten - je nach Priming Bedingung - Begriffe, die mit ORDM oder MRDM assoziiert sind, oder neutrale Begriffe. Im Beispielsatz bedeutet das, dass statt „interessante“ in der ORDM Bedingung „wenige“ und in der MRDM Bedingung „viele“ präsentiert wurde.

Nach dem Priming bearbeiteten die Probanden die 71 Durchgänge

3. Situative Einflussfaktoren

der Pferdewette, die in Kapitel 1.3.2.1 (S. 60) und Anhang A.2 (S. 171) beschrieben ist.

Dann bearbeiteten die Probanden den von Platzer (2008, Studie 2) entwickelten Manipulation-Check des Primings von Entscheidungsverhalten. Dazu wurden den Probanden in zufälliger Reihenfolge drei Szenarien vorgelegt, in denen eine Person eine Entscheidung trifft. Die Aufgabe der Probanden war es, Gründe für die Entscheidung zu benennen. So geht es z. B. in einem Szenario um einen Arzt, der sich für eine Behandlung zwischen zwei akuten Notfallpatienten, deren Symptome im Szenario beschrieben werden, für einen Patienten entscheidet; siehe Anhang C.3 (S. 190) für die drei verwendeten Szenarien. Die Probanden sollen also die Gründe einer Entscheidung, das sind die Cues für die Entscheidung, generieren. Die Hypothese ist, dass Probanden, die mit einer ORDM Entscheidungsverhalten geprimet werden, tendenziell auch weniger Gründe/Cues generieren als Probanden, die mit einer ORDM Entscheidungsverhalten geprimet werden.³

Dann wurden das *Bedürfnis nach kognitiver Geschlossenheit* (Schlink & Walther, 2007) und der *regulatorische Fokus* (Werth & Förster, 2007) am Computer erhoben; siehe Kapitel 4.2 (S. 135) für die Beschreibung der Skalen, Hypothesen des Zusammenhangs mit dem DLC und den Ergebnissen.

In der letzten Phase der Studie wurden demographische Daten der Probanden erhoben.

Scrambled Sentences Die 12 Priming Durchgänge, die ORDM, MRDM oder neutrale Begriffe beinhalteten, wurden in einer Studie von Platzer (2008, Studie 1, S. 45) entwickelt: Dazu wurden in Interviews Probanden gebeten, Begriffe, die diese mit einer in einer Vignette vorgestellten und erläuterten ORDM versus einer MRDM Strategie as-

³Das impliziert jedoch die Hypothese, dass Probanden nicht nur je nach Priming unterschiedlich viele Cues für eine *eigene* Entscheidung verwenden, sondern auch eine *fremde* Entscheidung als auf unterschiedlich vielen Cues basierend einschätzen; das Priming soll also nicht nur einen Einfluss auf das eigene Entscheidungsverhalten sondern auch auf die Wahrnehmung des Entscheidungsverhaltens von anderen Personen haben. Diese Hypothese wurde bisher empirisch nicht überprüft, die Ergebnisse in der Studie von Platzer (2008, S. 69) geben jedoch erste Evidenz, dass das Priming eines bestimmten Entscheidungsverhaltens einen Einfluss auf die Wahrnehmung von fremden Entscheidungsverhalten hat: Die Probanden der MRDM Priming Bedingung haben im Mittel über die drei Szenarien 6.17 (SD = 1.82) Gründe für eine Entscheidung generiert, die Probanden der ORDM Priming Bedingung haben 5.00 (SD = 1.57) Gründe generiert ($t(34) = 2.06, p < .05$).

soziieren, zu nennen.⁴ Die daraus entstandenen 65 Begriffe wurden um Begriffe, „die in der Literatur zur Heuristikforschung als eng mit einer Strategie in Verbindung stehend genannt werden“ (Platzer, 2008, S. 49), ergänzt und wiederum von Experten, die mit den Konzepten der Entscheidungstheorie vertraut sind, auf 22 MRDM und 18 ORDM Begriffe reduziert. Die Begriffe wurden dann abschließend von Probanden auf ihre Diskriminanz zwischen den beiden Strategieklassen und ihre Valenz bewertet und diesbezüglich - d. h. Begriffe mit einer hohen Diskriminanz und vergleichbaren Valenz zwischen den Bedingungen - auf die jeweils 12 Begriffe für die ORDM und MRDM Priming Bedingung reduziert, die auch in dieser Studie verwendet wurden.

Die restlichen acht Durchgänge der *Scrambled Sentences* Aufgabe stammen aus einer Studie von Walther et al. (2001) und dienten als Distraktoren, um das gemeinsame Thema der 12 Priming Durchgänge zu verbergen (Bargh & Chartrand, 2000); siehe Anhang C.2 (S. 188) für eine Auflistung aller in der Aufgabe verwendeten Scrambled Sentences.

Stichprobe Es nahmen 60 Probanden an der Studie teil. Davon waren 26 männlich und 34 weiblich. Das Alter lag zwischen 16 und 65 Jahren bei einem Mittelwert von 30.3 Jahren und einer Streuung von 12.6 Jahren.

3.1.1.2. Hypothesen

Probanden der Bedingung ORDM Priming haben ein geringeres DLC und Probanden der Bedingung MRDM Priming haben ein höheres DLC als die Probanden der neutralen Bedingung.

3.1.1.3. Ergebnisse

Das mittlere DLC liegt bei .786 bei einer Streuung von .101. Die Modelltestung gegen den Zufall⁵ ergibt, dass 41 von 60 (68%) Probanden als Anwender des DLC als Stoppregel der Informationssuche in den 20 Testdurchgängen der Pferdewette klassifiziert werden können. 3 Probanden können als Anwender einer ORDM-, 7 Probanden als Anwender einer MRDM- und 2 Probanden als Anwender einer Zufalls-Stoppregel

⁴In den Vignetten von Platzer (2008) wurden die *Take The Best* Heuristik und *Franklin's Rule* als Exemplare für eine ORDM und eine MRDM Strategie verwendet.

⁵siehe S. 64 in Kapitel 1.3.2.2 für Details zur Testung gegen den Zufall

3. Situative Einflussfaktoren

klassifiziert werden. 7 Probanden lassen sich nicht eindeutig nur einem der Modelle zuordnen.

Die Hypothese, dass Probanden der Bedingung ORDM Priming ein geringeres DLC und Probanden der Bedingung MRDM Priming (= Eperimentalbedingungen) ein höheres DLC als die Probanden der neutralen Bedingung (= Kontrollbedingung) haben, kann nicht bestätigt werden.

Beide Experimentalgruppen haben im Vergleich zur Kontrollgruppe ein niedrigeres DLC. Die Probanden der neutralen Gruppe haben ein DLC von .80 (SD = .09), die Probanden der ORDM Bedingung ein DLC von .75 (SD = .08) und die Probanden der MRDM Bedingung ein DLC von .79 (SD = 0.12).

In einem linearen Modell weichen die Prädiktorgewichte der dummykodierten Bedingungen nicht signifikant von der Kontrollgruppe ab; siehe Tabelle 3.1 (S. 98) für die Ergebnisse des multiplen linearen Modells.⁶ Die Bedingungen als Prädiktoren können 4% der Varianz im Kriterium, das ist das DLC, aufklären. Das lineare Modell unterscheidet sich nicht signifikant von einem Zufallsmodell ($F(2,57) = 1.30$, $p = .28$).

Tabelle 3.1. Multiple lineare Regression mit dem DLC als Kriterium und den experimentellen Bedingungen als Prädiktoren.

| Prädiktor | B | SEB | β | t | p |
|----------------|-------|------|---------|-------|-------|
| Intercept | .809 | .022 | - | 36.00 | < .05 |
| ORDM Bedingung | -.050 | .031 | -.411 | -1.58 | .11 |
| MRDM Bedingung | -.017 | .031 | -.142 | -0.55 | .58 |

Anmerkungen: $R^2 = .04$, $F(2,57) = 1.30$, $p = .28$. Die Bedingungen sind mit der neutralen Bedingung als Kontrollgruppe dummykodiert.

Die von den Probanden genannten Gründe für eine Entscheidung in den drei Szenarien wurde von zwei Kodierern ausgezählt. Die Inter-Rater Reliabilität liegt bei Cronbach's $\alpha = .97, .98, .96$ für Szenario 1, 2 und 3 (siehe Anhang C.3, S. 190), so dass die Kodierungen der Rater gemittelt wurden.

⁶Die Power liegt bei einer erwarteten mittleren Effektstärke von $f = .25$ und bei einem dreifach gestuften Faktor, 60 Probanden und einem $\alpha = .05$ bei $1-\beta = .37$.

Die gemittelten Kodierungen der Rater haben über die drei Szenarien hinweg eine interne Konsistenz von Cronbach's $\alpha = .85$, so dass für jeden Probanden ($n = 59$, da bei einem Probanden die Daten der Aufgabe nicht abgespeichert wurden) ein mittlerer Gesamtscore der genannten Gründe berechnet wurde.

Es zeigt sich, dass sich die Anzahl der genannten Gründe in den Experimentalgruppen nicht signifikant von der Kontrollgruppe unterscheiden. Im Mittel nennen die Probanden der neutralen Gruppe 1.87 ($SD = 1.14$) Gründe, die Probanden der ORDM Bedingung 1.83 ($SD = 1.24$) Gründe und die Probanden der MRDM Bedingung 1.70 ($SD = 0.49$) Gründe für die Fremdentcheidung in den Szenarien; siehe Tabelle 3.2 (S. 99) für die Ergebnisse des multiplen linearen Modells, in dem die Priming Bedingungen als Prädiktoren und die mittlere Anzahl der genannten Gründe für eine Fremdentcheidung als Kriterium eingehen.

Tabelle 3.2. Multiple lineare Regression mit der mittleren Anzahl an Gründen für eine Entscheidung als Kriterium und den experimentellen Bedingungen als Prädiktoren.

| Prädiktor | B | SEB | β | t | p |
|----------------|------|-----|---------|-------|-------|
| Intercept | 1.87 | .22 | - | 8.28 | < .05 |
| ORDM Bedingung | -.04 | .32 | -.03 | -0.12 | .89 |
| MRDM Bedingung | -.17 | .32 | -.14 | -0.54 | .58 |

Anmerkungen: $R^2 = .005$, $F(2,56) = 0.16$, $p = .85$. Die Bedingungen sind mit der neutralen Bedingung als Kontrollgruppe dummykodiert.

Die Priming Bedingungen als Prädiktoren können .5% der Varianz im Kriterium, das ist die mittlere Anzahl der genannten Gründe, aufklären. Das lineare Modell unterscheidet sich nicht signifikant von einem Zufallsmodell ($F(2,56) = 0.16$, $p = .85$).

Zusammenfassend zeigt sich in dieser Studie, dass das verwendete konzeptuelle Priming keinen Einfluss auf die Höhe des DLC hat. Die Befunde werden zusammenfassend nach der folgenden zweiten Priming Studie diskutiert.

3.1.2. Mindset Priming: Geschichten erfinden

Im Gegensatz zum konzeptuellen Priming, in dem eine mentale Repräsentation „passiv“ (Bargh & Chartrand, 2000, S. 265) durch semantisch assoziierte Begriffe aktiviert wird, wird beim Mindset Priming, das zum prozeduralen Priming gehört (Förster & Liberman, 2007, S. 215), eine bestimmte Art zu denken und/oder sich zu verhalten über eine Priming Aufgabe induziert, in der die Probanden „aktiv“ die interessierende „mentale Prozedur“ anwenden (Bargh & Chartrand, 2000, S. 265).⁷ Die Aktivierung der mentalen Prozedur zeigt dann einen carry-over Effekt (Bargh & Chartrand, 2000, S. 265) auf eine nachfolgende Zielaufgabe.

Es gibt keine etablierten Methoden zum Mindset Priming, so dass die Art des Primings, je nachdem welche mentale Prozedur aktiviert werden soll, über Studien hinweg unterschiedlich ausfällt (Gollwitzer, Heckhausen & Steller, 1990; Higgins & Chaires, 1980; Rim, Uleman & Trope, 2009; Sassenberg & Moskowitz, 2005; Sassenberg, Moskowitz, Jacoby & Hansen, 2007).

So sollten z. B. Probanden in einer Studie von Friedman, Fishbach, Förster und Werth (2003) in der Priming Aufgabe - einer perzeptuellen Suchaufgabe - so schnell wie möglich angeben, ob eine bestimmte Zahl in einem Feld von Zahlen vorkommt. Manipuliert wurde die Distanz der Positionen zwischen den Zahlen auf dem Bildschirm, so dass eine geringe Distanz vs. große Distanz ein perzeptuell enges vs. weites Suchfeld für die Probanden bedeutete. In der folgenden Zielaufgabe - einer Aufgabe zur Messung von Kreativität - sollten die Probanden kreative Verwendungsmöglichkeiten für einen Ziegelstein angeben. Die Probanden, die auf einen engen „Fokus“ geprimet wurden, produzierten weniger kreative Lösungen als Probanden, die auf einen weiten Fokus geprimet wurden.

In der vorliegenden Studie soll über ein Mindset Priming ORDM vs. MRDM angeregt werden: In einer Priming Aufgabe sollten Probanden *eine* vs. *alle* Information verwenden, was zu einem carry-over Effekt zu ORDM vs. MRDM in der nachfolgenden DLC Pferdewette, d. h. der

⁷siehe aber Stapel und Koomen (2001, Fußnote 2, S. 136): „The distinction between mindset and knowledge priming is somewhat artificial. One may argue that to a certain extent, all priming techniques activate knowledge of some kind or another.“ und siehe Förster und Liberman (2007, S. 216): „[I]t seems that the activation of semantic and procedural knowledge may occur simultaneously and might have independent effects on information processing. [...] Further research is needed to distinguish between the two kinds of priming.“

Zielaufgabe des Primings, führen sollte.

3.1.2.1. Methoden

Die Studie bestand aus vier Phasen. In der ersten Phase wurde das DLC gemessen. Dann bearbeiteten die Probanden je nach Bedingung eine Priming Aufgabe, die ORDM oder MRDM Entscheidungsverhalten anregen sollte. Dann erfolgte eine zweite Messung des DLC und abschließend eine Erhebung von Persönlichkeitsvariablen. Im Folgenden werden die Aufgaben der einzelnen Phasen genauer beschrieben.

Die Probanden wurden schriftlich über ihre Rechte als Teilnehmer an einer psychologischen Studie aufgeklärt. Die Durchführung aller Aufgaben fand am Computer statt. Die Probanden erhielten für ihre Teilnahme für ihr Studium notwendige Versuchspersonenstunden gutgeschrieben.

Die Probanden bearbeiteten zuerst die 71 Durchgänge der Pferdewette. In den ersten 51 Durchgängen, der *Messphase* des DLC, wurde das DLC bestimmt, in den folgenden 20 Durchgängen, der *Testphase* des DLC, wurde das gemessene DLC adaptiv getestet; siehe Kapitel 1.3.2.1 (S. 60) und Anhang A.2 (S. 171) für eine genaue Beschreibung der Aufgabe und des Messmodells.

Die dann nachfolgende Priming Aufgabe wurde als Kreativitäts-Aufgabe getarnt. Die Probanden sollten in der Aufgabe vier kurze Geschichten schreiben. Dazu wurden für jede Geschichte acht Begriffe - wie z. B. von konkreten Objekten wie einer *Erdbeere* bis zu abstrakten Konzepten wie *Ehrenhaftigkeit* - angezeigt. Die acht Begriffe wurden per Zufallsposition auf einer 3×3 Matrix auf dem Bildschirm präsentiert.

In der Priming Bedingung, in der ORDM induziert werden sollte, wurde per Zufall *einer* der acht Begriffe rot markiert. Die Probanden wurden in der Instruktion angewiesen, dass die Geschichte den *einen* rot markierten Begriff enthalten muss und keinen der übrigen sieben Begriffe enthalten darf. Die Instruktion lautete:

Es gibt nur eine Regel beim Schreiben Ihrer Geschichte zu beachten: Bitte verwenden Sie nur den rot markierten Begriff und NICHT die anderen Begriffe in Ihrer Geschichte, d. h. nur der rot markierte Begriff und nicht die anderen Begriffe sollten in Ihrer Geschichte vorkommen!

In der Priming Bedingung, in der MRDM induziert werden sollte, war *keiner* der acht Begriffe rot markiert. Die Probanden wurden in der

3. Situative Einflussfaktoren

Instruktion angewiesen, dass die Geschichte *alle* acht Begriffe enthalten muss. Die Instruktion lautete:

Es gibt nur eine Regel beim Schreiben Ihrer Geschichte zu beachten: Bitte verwenden Sie alle 8 aufgelisteten Begriffe in Ihrer Geschichte, d. h. alle 8 Begriffe sollten in Ihrer Geschichte vorkommen!

Die ORDM Instruktion sollte das Mindset anregen, nur einen von vielen Elementen/Informationen zur Lösung einer Aufgabe (= Ziel) zu verwenden, während die MRDM Instruktion das Mindset angeregt werden sollte, alle Elemente/Informationen zur Lösung einer Aufgabe (= Ziel) zu verwenden. Bis auf den Unterschied in der Instruktion und die rote Markierung des Wortes in der ORDM Bedingung, waren die Priming Bedingungen identisch.

Die Geschichten sollten auf ein Blatt Papier geschrieben werden, während auf dem Bildschirm die Wörter und die verbleibende Zeit für das Schreiben der Geschichte eingeblendet wurden. Die Probanden hatten für jede Geschichte vier Minuten Zeit; nach Ablauf der Zeit wurden die Wörter am Computer durch eine rote aufblinkende Maske verdeckt und die Probanden aufgefordert, das Schreiben der Geschichte zu beenden und einen Button anzuklicken, um die nächsten acht Wörter für die folgende Geschichte zu erhalten.

Die Begriffe, die für die vier Durchgänge verwendet wurden, stammen aus einem Set von Begriffen, die in einer Studie von Baschek, Bredenkamp, Ohrle und Wippich (1977) - im Folgenden zitiert nach Baschek, Bredenkamp, Ohrle und Wippich (1994) - auf ihre Bildhaftigkeit, Konkretheit-Abstraktheit und Bedeutungshaltigkeit normiert wurden.

Bildhaftigkeit bedeutet, wie schnell ein Begriff „das Auftreten einer bildhaften Vorstellung oder eines Klangs“ (Baschek et al., 1994, S. 177) bewirkt. Konkretheit-Abstraktheit bedeutet, wie stark sich ein Begriff auf „sinnlich erfahrbare Merkmale von Dingen oder Personen bezieht“ (Baschek et al., 1994, S. 178). Bedeutungshaltigkeit bedeutet, „in welchem Ausmaß sie [d. h. die Begriffe] andere Wörter hervorrufen“ (Baschek et al., 1994, S. 180); Baschek et al. (1994, S. 180) umschreiben Bildhaftigkeit auch als das Ausmaß der „Assoziationen“ eines Begriffs mit anderen Wörtern. Die Skala der Normierungsstudie war siebenfach gestuft, wobei hohe Werte hohe Bildhaftigkeit, hohe Konkretheit und hohe Bedeutungshaltigkeit bedeuten. Die Wörter wurden zum Vergleich zu anderen Normierungsstudien in eine Skala von -20 bis +20 -

inklusive dem Nullpunkt - transformiert.

Die Mittelwerte und Streuungen der acht Begriffe pro Geschichte auf den drei Dimensionen sind in Tabelle 3.3 (S. 103) abgebildet. Alle 32 verwendeten Begriffe und deren Normwerte befinden sich im Anhang C.4 (S. 191).

Tabelle 3.3. Mittelwert und Streuung für die Bildhaftigkeit, Konkretheit und Bedeutungshaltigkeit gemittelt über die acht Wörter für das jeweilige Set (Baschek et al., 1994).

| Begriffe | Bildhaftigkeit | | Konkretheit | | Bedeutungshaltigkeit | |
|----------|----------------|------|-------------|------|----------------------|------|
| | M | SD | M | SD | M | SD |
| Set 1 | 16.66 | 0.95 | 12.98 | 4.31 | 12.45 | 0.80 |
| Set 2 | 15.46 | 1.83 | 15.80 | 2.73 | 6.32 | 0.87 |
| Set 3 | -1.43 | 1.87 | -3.32 | 2.23 | -7.67 | 0 |
| Set 4 | -5.51 | 5.85 | -7.00 | 3.26 | -13.10 | 1.53 |

Die Begriffe wurden so ausgewählt, dass die vier Sets von Begriffen, die per Zufallsreihenfolge präsentiert wurden, unterschiedlich stark bedeutungshaltig sind, um es für die Probanden *beider* Priming Bedingungen über die Sets hinweg einfacher bzw. schwieriger zu machen, Geschichten zu erfinden. Das heißt, wenn die Begriffe starke vs. geringe Assoziationen mit anderen Wörtern hervorrufen, dann ist die Wahrscheinlichkeit, dass die Begriffe eines Sets untereinander assoziiert sind, höher vs. geringer. Eine *hohe* Assoziation der Begriffe eines Sets sollte es für die Probanden der ORDM vs. MRDM Bedingung erschweren vs. erleichtern, eine Geschichte zu erfinden. Eine *niedrige* Assoziation der Begriffe eines Sets sollte es für die Probanden der ORDM versus MRDM Bedingung erleichtern vs. erschweren, eine Geschichte zu erfinden. Durch die Variation der Bildhaftigkeit der Begriffe wird sichergestellt, dass die Aufgabe - das Erfinden von Geschichten - für die Priming Bedingungen nicht unterschiedlich schwierig ist.

Um den Einfluss des Primings auf das DLC zu messen, bearbeiteten die Probanden dann noch einmal Durchgänge der Pferdewette. Es wurden nicht noch einmal alle 51 Durchgänge der Messphase präsentiert, um ein DLC nach dem Priming - das DLC Priming - zu messen, sondern *nur* noch einmal die 20 Aufgaben der Testphase des DLC präsentiert. Über das Antwortverhalten über die 20 Durchgänge wurde dann der

3. Situative Einflussfaktoren

DLC Algorithmus zur Berechnung des DLC Priming angewendet (s. o.).⁸

Dann wurde das *Bedürfnis nach kognitiver Geschlossenheit* (Schlink & Walther, 2007) am Computer erhoben; siehe Kapitel 4.2 (S. 135) für die Beschreibung der Skala, die Hypothese des Zusammenhangs mit dem DLC und dem Ergebnis.

In der letzten Phase der Studie wurden demographische Daten der Probanden erhoben.

Stichprobe Es nahmen 39 Probanden an der Studie teil. Davon waren 3 männlich und 36 weiblich. Das Alter lag zwischen 19 und 41 Jahren bei einem Mittelwert von 24.0 Jahren und einer Streuung von 5.6 Jahren.

3.1.2.2. Hypothesen

Probanden der ORDM Priming Bedingung haben ein geringeres DLC Priming als Probanden der MRDM Priming Bedingung.

3.1.2.3. Ergebnisse

Das mittlere DLC liegt bei .778 bei einer Streuung von .121. Die Modelltestung gegen den Zufall⁹ ergibt, dass 24 von 39 (61%) Probanden als Anwender des DLC als Stoppregel der Informationssuche in den 20 Testdurchgängen der Pferdewette klassifiziert werden können. 3 Probanden können als Anwender einer ORDM-, 5 Probanden als Anwender einer MRDM- und 2 Probanden als Anwender einer Zufalls-Stoppregel klassifiziert werden. 5 Probanden lassen sich nicht eindeutig nur einem der Modelle zuordnen.

Die Hypothese zu dem Unterschied im DLC Priming zwischen den beiden Bedingungen kann nicht bestätigt werden. Es ist nicht der Fall, dass Probanden, die aus einem der Begriffe eine Geschichte erfinden sollten (ORDM), ein signifikant geringeres DLC Prime haben als Probanden, die aus allen Begriffen eine Geschichte erfinden sollten (MRDM).

Der Mittelwert der Probanden der ORDM Bedingung liegt bei .812 und einer Streuung von .144. Der Mittelwert der Probanden der MRDM Bedingung liegt bei .783 und einer Streuung von .134. Der Mittelwertsunterschied ist mit $t(37) = 0.66$, $p = .74$ und einem Cohen's $d = .21$

⁸Es wurde aus zeitlichen Gründen nicht noch einmal alle 51 Aufgaben der Testphase präsentiert.

⁹siehe S. 64 in Kapitel 1.3.2.2 für Details zur Testung gegen den Zufall

nicht signifikant und zeigt deskriptiv in die hypothesen*inkonforme* Richtung.¹⁰

Es ist wahrscheinlich, dass es durch die natürliche Obergrenze des DLC von 1 schwieriger bzw. leichter ist, durch eine situationale Manipulation einen schon hohen Ausgangswert im DLC noch weiter zu erhöhen bzw. zu erniedrigen und dass es durch die Mess-Untergrenze des DLC von .5 schwieriger bzw. leichter ist, durch eine situationale Manipulation einen schon niedrigen Ausgangswert im DLC noch weiter zu erniedrigen bzw. zu erhöhen. Es wurde unter Anwendung des Messalgorithmus des DLC (s. o.) ein DLC über die 20 Aufgaben der Testphase berechnet. Um den deskriptiven Unterschied im DLC der Testphase und somit für die uninteressanten interindividuellen Unterschiede im DLC *vor* dem Priming zu kontrollieren, wurde dann eine multiple lineare Regression mit der Priming Bedingung und dem DLC der Testphase als Prädiktoren und dem DLC Priming als Kriterium gerechnet; siehe Tabelle 3.4.

Tabelle 3.4. Multiple lineare Regression mit DLC Priming als Kriterium und der experimentellen Bedingung und dem DLC der Testphase als Prädiktoren.

| Prädiktor | B | SEB | β | t | p |
|------------------------|------|------|---------|-------|-------|
| Intercept | .074 | .052 | - | 1.42 | .16 |
| Bedingung MRDM Priming | .001 | .017 | .006 | 0.10 | .92 |
| DLC Testphase | .915 | .063 | .925 | 14.47 | < .05 |

Anmerkungen: $R^2 = .85$, $F(2,36) = 106.2$, $p < .05$. Die Bedingungen sind mit der Bedingung ORDM Priming als Vergleichsgruppe dummykodiert.

Die Prädiktoren können 85% der Varianz des DLC Priming aufklären. Die Varianzaufklärung geht aber auf den Prädiktor des DLC der Testphase zurück; das Gewicht des Prädiktors für die Priming Bedingung ist mit einem $t = 0.10$ ($p = .92$) nicht signifikant. Die Bedingung kann über das DLC der Messphase hinaus *nur* .0001% einzigartige Varianz aufklären, was sich im Vergleich des Gesamtmodells mit einem um die Priming Bedingung reduzierten linearen Modell zeigt.

¹⁰Die Power liegt bei einer erwarteten mittleren Effektstärke von Cohen's $d = .5$ und einer Stichprobengröße von 19 für die ORDM resp. 20 für die MRDM Bedingung und einem $\alpha = .05$ bei $1 - \beta = .45$.

3. Situative Einflussfaktoren

Zusammenfassend zeigt sich in dieser Studie, dass das verwendete Mindset Priming keinen Einfluss auf die Höhe des DLC hat.

3.1.3. Diskussion

Die Höhe des DLC lässt sich nicht durch das in den Studien verwendete konzeptuelle Priming und Mindset Priming beeinflussen.

Das kann *methodische* Gründe haben: Es ist möglich, dass die Manipulationen nicht gewirkt haben. Ein Hinweis dafür gibt die Aufgabe des Manipulation-Checks beim konzeptuellen Priming: Die Probanden generierten nach einem konzeptuellen Priming von MRDM nicht mehr Gründe - wie das in der Studie von Platzer (2008) der Fall war - als Probanden nach einem Priming von ORDM.

Es zeigte sich weiterhin bei der Aufgabe des Manipulation-Checks, dass die Anzahl der genannten Gründe im Mittel über alle Aufgaben und Probanden hinweg mit 1.80 Gründen bei einer Streuung von 0.99 Gründen gering ausfiel.¹¹ Das kann auf eine geringe Motivation der Probanden hindeuten, was problematisch für alle Messvariablen der Studie ist.

Die verwendete Methode des Mindset Primings wurde für diese Studie entwickelt und ist nicht bewährt/validiert, so dass es möglich ist, dass die Methode zu unspezifisch ist, um ORDM vs. MRDM hervorzurufen. Es ist weiterhin möglich, dass das reduzierte Aufgabenset zur Messung des DLC nach dem Priming nicht ausreichend war, um Veränderungen im DLC reliabel zu messen.

Es besteht weiterhin in beiden Studien ein *statistisches* Problem: Die Power bei der Anzahl der erhobenen Probanden von $n = 60$ bei drei Bedingungen für das konzeptuelle Priming und $n = 39$ bei zwei Bedingungen für das Mindset Priming, einer mittleren erwarteten Effektstärke und einem α -Fehler von .05 fällt mit einem Wert von .37 für das konzeptuelle Priming und .45 für das Mindset Priming gering aus, so dass die Wahrscheinlichkeit für einen β -Fehler in beiden Studien hoch ist.

Es besteht in beiden Studien ein *konzeptuelles* Problem: Wenn in der DLC Pferdewette-Aufgabe tatsächlich ein psychologisch bedeutsames Konstrukt - nämlich das Anspruchsniveau an gewünschter Urteilssicherheit (Hausmann-Thürig, 2004) - gemessen wird, dann führt ein spezifisches Anspruchsniveau an gewünschter Urteilssicherheit dazu,

¹¹ zur Erinnerung: In der Studie von Platzer (2008) lag der Mittelwert bei 5.58 Gründen.

dass vor einer Entscheidung mehr bzw. weniger Cues konsultiert werden. ORDM vs. MRDM Stoppverhalten ist dann also die *Konsequenz* einer *ursächlich* zugrundeliegenden Ausprägung des DLC Konstruktes.

Wenn in einem konzeptuellen Priming oder Mindset Priming *nur* die Konsequenz einer Konstruktausprägung induziert wird, bedeutet das nicht zwangsläufig, dass das Priming auf die Ausprägung des Konstruktes wirkt bzw. stark wirkt. Dies wird besonders im Mindset Priming deutlich: Durch die Hypothese, dass die Verwendung einer bzw. aller Informationen in einer Priming Aufgabe dazu führt, dass die Probanden in der DLC Aufgabe wenige vs. viele Cues aufdecken, wird das DLC zu einem Messinstrument von ORDM und MRDM vs. einem psychologisch bedeutsamen Konstrukt „degradiert“. Selbst wenn das Priming eine Tendenz zu ORDM vs. MRDM induziert, kann diese Verhaltens-tendenz durch das individuelle Anspruchsniveau an gewünschter Urteilssicherheit, das - wie die Studie zur Validität des DLC nahelegt (Kapitel 2.2, S. 79) - zeitlich stabil ist und dessen Ausprägung durch das Priming nicht direkt verändert wurde, überschrieben werden.

Weiterhin zeigt sich beim Mindest Priming auch ein *konzeptuelles* Problem: So mussten die Probanden in *beiden* Priming Bedingungen (ORDM und MRDM) alle dargebotenen Begriffe berücksichtigen. Es war nur der Umgang mit den Begriffen unterschiedlich, das heißt Verwendung eines Begriffs (und damit Ausblenden aller anderen Begriffe) in der ORDM vs. Verwendung aller Begriffe in der MRDM Bedingung. Hier wäre es sinnvoller gewesen, nur einen (ORDM) vs. viele (MRDM) Begriffe als Stimuli darzubieten.

Hier wird auch ein weiteres konzeptuelles Problem deutlich: Die Ergebnisse der Studie zur Konstruktvalidität des DLC (siehe Studie 2, Kapitel 2.2, S. 79) legen nahe, dass die Art der Cue-Integration nicht vom DLC erfasst wird, das DLC also *nur* den Abbruch der Suche nach Cues abbilden kann: Ein hohes DLC bedeutet nicht zwangsläufig eine kompensatorische Integration von Cues für eine Entscheidung. Es ist somit möglich, dass das verwendete konzeptuelle Priming, das auch Begriffe zur Art der Integration von Informationen beinhaltet (z. B. komplex), und das verwendete Mindset Priming, das auf die Integration von Cues abzielt, nicht spezifisch für das DLC sind.¹²

Es ist also für eine Folgestudie sinnvoll, das Priming so zu gestalten, dass es eine spezifische Ausprägung des angenommenen zugrun-

¹²Diese Überlegung war zum Zeitpunkt der Durchführung der Priming Studien jedoch noch nicht möglich, da die Studie zur Konstruktvalidität des DLC chronologisch nach den Priming Studien durchgeführt wurde.

3. Situative Einflussfaktoren

deliegenden Konstruktes - das ist das Anspruchsniveau an gewünschter Urteilsicherheit - bewirkt. So gab es *nur* einen konzeptuell zum DLC verwandten Satz in der ORDM Bedingung der Priming Aufgabe, das ist: Peter, Risiken, muss, eingehen, man. Letztendlich ist es jedoch auch möglich, dass sich Entscheidungsverhalten in probabilistischen Entscheidungsaufgaben im Kontext der adaptiven Werkzeugkiste (Gigerenzer et al., 1999) nicht durch Priming beeinflussen lässt, für diese Schlussfolgerung reichen jedoch die zwei hier berichteten ersten Studien zum Thema nicht aus.

3.2. Need for Cognitive Closure

Need for Cognitive Closure (Kruglanski & Webster, 1996; Kruglanski, 2004) bzw. das *Bedürfnis nach kognitiver Geschlossenheit* (Schlink & Walther, 2007) - im Folgenden mit NFCC abgekürzt - ist eine motivationale Tendenz im Prozess der Generierung von Wissen: Menschen mit einem hohen NFCC haben das *Bedürfnis*, im Prozess der Wissensbildung schnell zu einer Schlussfolgerung zu kommen, um den aversiv erlebten Zustand der Ambiguität der Situation, in der es noch keine definitive Antwort auf eine Frage gibt, zu verlassen, und eine einmal getroffene Schlussfolgerung nicht mehr zu verändern, um nicht in den aversiv erlebten Zustand der Ambiguität zurückzufallen.¹³

Die Tendenz, eine Schlussfolgerung schnell treffen zu wollen, nennt sich *seizing*; die Tendenz, eine dauerhafte Schlussfolgerung treffen zu wollen, nennt sich *freezing* (Kruglanski & Webster, 1996).¹⁴

Ein hohes NFCC kann chronisch sein und über einen Fragebogen (Schlink & Walther, 2007) gemessen werden oder situational durch externe Bedingungen, wie z. B. Zeitdruck, Lärm und Hitze, oder durch internale Bedingungen, wie z. B. Müdigkeit, induziert werden (Kruglanski, 2004).

In einer Studie von Kruglanski, Webster und Klem (1993) sollten Probanden über einen Rechtsfall mit einem Diskussionspartner entscheiden. Die Hälfte der Probanden erhielt vor der Diskussion mit dem Partner Informationen zum Fall und konnte sich so vorab eine Meinung bilden. Probanden mit einer hohen Ausprägung auf der NFCC Skala und

¹³bei Kruglanski und Webster (1996, S. 264) heißt das: „The need for cognitive closure refers to individuals’ desire for a firm answer to a question and an aversion toward ambiguity.“

¹⁴Kruglanski und Webster (1996, S. 265) nennen *seizing* und *freezing* auch *urgency* und *permanence* Tendenz.

Probanden, denen situational hohes NFCC über Lärm in der Erhebungssituation induziert wurde, zeigten die Tendenz, möglichst schnell mit dem Partner übereinzustimmen.

Wenn *keine* Vorabinformation vorlag, äußerte sich die Tendenz zu einem schnellen Konsensus darin, dass die Probanden einen Partner bevorzugten, der überzeugend ist; eine Konsequenz der *seizing* Tendenz, die *vor* einer Schlussfolgerung eintritt. Wenn Vorabinformation vorlag, äußerte sich die Tendenz zu einem schnellen Konsensus darin, dass die Probanden einen Partner bevorzugten, der sich leicht überzeugen lässt; eine Konsequenz der *freezing* Tendenz, die *nach* einer einmal getroffenen Schlussfolgerung eintritt.

Die Forschung zum NFCC verbindet ein breites Themenspektrum aus der klassischen Sozialpsychologie: So lassen sich z. B. NFCC Effekte auf die Bildung von Stereotypen und die Stärke von Ankereffekten (Kruglanski & Freund, 1983), auf die soziale Beeinflussung (Kruglanski et al., 1993), auf den *correspondence bias* aus der Attributionsforschung (Webster, 1993), auf den *linguistic intergroup bias* aus der psycholinguistischen Stereotypenforschung (Webster, Kruglanski & Pattison, 1997), auf den *ingroup bias* aus der gruppenmotivationalen Stereotypenforschung (Shah, Kruglanski & Thompson, 1998) und auf die Entstehung von (politisch konservativen) Einstellungen (Jost, Glaser, Kruglanski & Sulloway, 2003) nachweisen.

In den folgenden Studien soll der situationale Einfluss des NFCC auf die Höhe des DLC untersucht werden. Ein hohes situationales NFCC führt *vor* einer Entscheidung für eine Option - das ist in der Messaufgabe des DLC ein Pferd - zu *seizing* Tendenzen: Die Probanden haben das Bedürfnis, schnell zu einer Lösung zu kommen, was dazu führt, dass tendenziell weniger Cues in der Pferdewette aufgedeckt werden und somit das DLC geringer ausfällt.

3.2.1. Aufgabenattraktivität: Videos anschauen

Eine attraktive vs. unattraktive Folgeaufgabe kann hohes vs. niedriges NFCC hervorrufen (Mayseless & Kruglanski, 1987; Webster, 1993). Eine attraktive vs. unattraktive Folgeaufgabe führt dazu, dass die aktuelle Aufgabe tendenziell schneller vs. langsamer beendet und Schlussfolgerungen auf weniger vs. mehr Informationen getroffen werden (Webster, 1993), was eine Konsequenz der *seizing* Tendenz bei hohem NFCC ist. Ein hoher (situational induzierter) NFCC Wert sollte in einer geringeren Informationssuche vor einer Entscheidung resultieren und somit zu

3. Situative Einflussfaktoren

einem geringeren DLC führen.

3.2.1.1. Methoden

Die Studie bestand aus fünf Phasen. In der ersten Phase wurde ein hohes vs. niedriges NFCC über eine Instruktion induziert. In der zweiten Phase folgte die DLC Messaufgabe. Dann folgte eine Filler Aufgabe, um in der dritten Phase erneut das DLC adaptiv zu testen. In der vierten Phase wurden Persönlichkeitsvariablen erhoben. Abschließend konnten die Probanden, je nach situationaler NFCC Bedingung, unterschiedliche kurze Videos anschauen. Im Folgenden werden die Aufgaben der einzelnen Phasen genauer beschrieben.

Die Probanden wurden schriftlich über ihre Rechte als Teilnehmer an einer psychologischen Studie aufgeklärt. Die Durchführung aller Aufgaben fand am Computer statt. Die Probanden erhielten für ihre Teilnahme für ihr Studium notwendige Teilnahmestunden gutgeschrieben und die drei Probanden mit der höchsten Punktzahl erhielten zusätzlich einen Kinogutschein.

Den Probanden wurde in der Instruktion der Studie mitgeteilt, dass die Studie aus 2 Phasen besteht.

In der Bedingung hohes NFCC wurden den Probanden mitgeteilt, dass sie nach dem Bearbeiten der Pferdewette-Aufgabe und einer Lotteripräferenzaufgabe in der ersten Phase der Studie, in der zweiten Phase der Studie fünf *Comedy* Videos anschauen und auf ihre Unterhaltsamkeit hin bewerten werden. Die Titel der fünf Videos wurde angezeigt, darunter war z. B. ein Video mit dem Titel „Harald Schmidt spricht über Heavy Metal Musik“ und ein Video mit dem Titel „Helge Schneider interviewt sich selbst“.

In der Bedingung niedriges NFCC wurden den Probanden mitgeteilt, dass sie nach dem Bearbeiten der Pferdewette-Aufgabe und einer Lotteripräferenzaufgabe in der ersten Phase der Studie, in der zweiten Phase der Studie fünf *Statistik* Videos anschauen und zu den Inhalten der Statistik Videos befragt werden. Die Titel der fünf Videos wurde angezeigt, darunter war z. B. ein Video mit dem Titel „Testtheorie: True Score Modell“ und ein Video mit dem Titel „Skalentypen“.

Den Probanden wurden in beiden Bedingungen mitgeteilt, dass die Studie zeitlich begrenzt ist und die Anzahl der Videos, die die Probanden am Ende anschauen werden, davon abhängig ist, wie zügig sie die Aufgaben der ersten Phase der Studie bearbeiten; siehe Anhang B.4 (S. 178) für die genauen Instruktionen der Studie. Weiterhin wurde

den Probanden über die gesamte Studie hinweg bei den Instruktionsschirmen zu jeder Aufgabe auf einer Zeitleiste angezeigt, wie viele Aufgaben sie schon bearbeitet haben und wie viele Aufgaben sie noch bearbeiten müssen. Somit wurde die Distanz zur Video Aufgabe der Phase 2 über die Studie hinweg wiederholt salient gemacht.

Dann bearbeiteten die Probanden die 51 Messdurchgänge der Pferdewette, die in Kapitel 1.3.2.1 (S. 60) und Anhang A.2 (S. 171) beschrieben ist.

Dann bearbeiteten die Probanden Lotteripräferenzaufgaben, deren Inhalt für die vorliegende Studie nicht relevant ist, es aber ermöglichte, die erste Phase der Studie zu verlängern. Die Probanden benötigten ca. 10 Minuten zu Bearbeitung der Präferenzaufgabe.

Dann bearbeiteten die Probanden die 20 Durchgänge der Testphase der Pferdewette.

Dann wurde die deutsche Version der *Positive and Negative Affect Schedule (PANAS)* (Krohne, Egloff, Kohlmann & Tausch, 1996) erhoben. Über den PANAS können in jeweils 10 Fragen der momentane positive und der negative Affekt einer Person erhoben werden. Die Probanden sollten dazu auf einer fünfstufigen Skala angeben, wie z. B. freudig erregt sie sich momentan fühlen bzw. wie z. B. gereizt sie sich momentan fühlen. Die beiden Subskalen weisen in der Studie von Krohne et al. (1996, Tabelle 3, S. 145) mit einem Cronbach's α von .85 für positiven Affekt und einem Cronbach's α von .86 für negativen Affekt eine gute interne Konsistenz auf.

Dann wurden das *Bedürfnis nach kognitiver Geschlossenheit* (Schlink & Walther, 2007) und der *regulatorische Fokus* (Werth & Förster, 2007) am Computer erhoben; siehe Kapitel 4.2 (S. 135) für die Beschreibung der Skalen, Hypothesen des Zusammenhangs mit dem DLC und den Ergebnissen.

Dann sollten die Probanden die Videos anschauen - insgesamt sollten sie drei von fünf Comedy bzw. Statistik Videos auswählen - und auf ihre Attraktivität hin bewerten bzw. den Inhalt des Videos nach jedem Video zusammenfassen.

Im letzten Abschnitt der Studie wurden demographische Daten der Probanden erhoben.

Stichprobe Es nahmen 40 Probanden an der Studie teil. Davon waren 10 männlich und 30 weiblich. Das Alter lag zwischen 19 und 44 Jahren bei einem Mittelwert von 23.8 Jahren und einer Streuung von 5.7 Jahren.

3. Situative Einflussfaktoren

3.2.1.2. Hypothesen

Probanden, die eine attraktive Folgeaufgabe bearbeiten sollen (Bedingung hohes NFCC), haben im Vergleich zu Probanden, die eine unattraktive Folgeaufgabe bearbeiten sollen (Bedingung niedriges NFCC), ein niedrigeres DLC.

3.2.1.3. Ergebnisse

Das mittlere DLC liegt bei .755 bei einer Streuung von .113. Die Modelltestung gegen den Zufall¹⁵ ergibt, dass von 31 von 40 (77.5%) Probanden als Anwender des DLC als Stoppregel der Informationssuche in den 20 Testdurchgängen der Pferdewette klassifiziert werden können. 2 Probanden können als Anwender einer ORDM-, 1 Proband als Anwender einer MRDM- und 1 Proband als Anwender einer Zufalls-Stoppregel klassifiziert werden. 5 Probanden lassen sich nicht eindeutig nur einem der Modelle zuordnen.

Die Probanden in der Bedingung hohes NFCC haben kein signifikant niedrigeres DLC als die Probanden in der Bedingung niedrig NFCC. Der Mittelwert liegt für das DLC der Messphase für die Bedingung hohes NFCC bei .753 und einer Streuung von .173, der Mittelwert liegt für das DLC der Messphase für die Bedingung niedriges NFCC bei .758 und einer Streuung von .120. Der Unterschied ist mit einem $t(38) = -0.12$, $p = .45$ und Cohen's $d = -0.03$ nicht signifikant.¹⁶

Die Probanden bearbeiten die Durchgänge der DLC Testphase nach den Aufgaben zu den Präferenzentscheidungen. Es ist möglich, dass sich die NFCC Bedingung erst über die Studie kurz vor der attraktiven vs. wenig attraktiven Folgeaufgabe zeigt. Über die 20 Aufgaben der Testphase der Pferdewette kann der gleiche DLC Algorithmus angewendet werden, so dass ein zweites DLC berechnet werden kann. Das DLC der Testphase unterscheidet sich jedoch in den NFCC Gruppen auch nicht signifikant voneinander. Die Probanden in der Bedingung hohes NFCC haben kein signifikant niedrigeres DLC der Testphase als die Probanden in der Bedingung hohes NFCC. Der Mittelwert liegt für das DLC der Testphase für die Bedingung hohes NFCC bei .736 und einer Streuung von .173, der Mittelwert liegt für das DLC der Testphase für die Bedingung niedriges NFCC bei .747 und einer Streuung von .144. Der

¹⁵siehe S. 64 in Kapitel 1.3.2.2 für Details zur Testung gegen den Zufall

¹⁶Die Power liegt bei einer erwarteten mittleren Effektstärke von Cohen's $d = .5$, 40 Probanden und einem $\alpha = .05$ bei $1 - \beta = .46$.

Unterschied ist mit einem $t(38) = -0.21$, $p = .41$ und Cohen's $d = -0.06$ nicht signifikant.

Die Skala positiver Affekt und negativer Effekt des PANAS (Krohne et al., 1996) zeigt mit einem Cronbach's α von .88 und .87 eine hohe interne Konsistenz.

Die Probanden der Bedingung hohes NFCC vs. niedriges NFCC zeigen mit einem Mittelwert von 30.2 vs. 26.25 und einer Streuung von 6.01 vs. 7.29 einen erhöhten positiven Affekt. Der Unterschied ist mit einem $t = 1.86$, $p < .05$, Cohen's $d = 0.59$ signifikant.¹⁷

Die Probanden der Bedingung hohes NFCC vs. niedriges NFCC zeigen mit einem Mittelwert von 11.7 vs. 14.9 und einer Streuung von 1.7 vs. 5.8 einen erniedrigten negativen Affekt. Der Unterschied ist mit einem $t = -2.39$, $p < .05$, Cohen's $d = -0.75$ signifikant.

Es zeigt sich also, dass sich das DLC nicht durch eine Manipulation des NFCC beeinflussen lässt. Die Ergebnisse werden nach der folgenden Studie, in der eine weitere Induktion des NFCC als situationale Manipulation des DLC getestet wird, zusammengefasst diskutiert.

3.2.2. Erhebungssituation: Lärm ertragen

Lärm in der Erhebungssituation ist eine bewährte situationale Manipulation eines hohen NFCC (Kruglanski & Webster, 1991, 1996; Webster et al., 1997).¹⁸ Lärm wirkt aversiv, die Tendenz, zu einer schnellen und dauerhaften Schlussfolgerung in einem Prozess der Wissensgenerierung zu kommen, ist erhöht.

3.2.2.1. Methoden

Die Studie bestand aus vier Phasen. In der ersten Phase wurde das DLC gemessen. Über das Einspielen vs. Fehlen eines unangenehmen Tons wurde situational ein hohes vs. niedriges NFCC induziert. In der zweiten Phase wurde die Wirkung der situationalen Manipulation des NFCC in einer bewährten Aufgabe zum NFCC überprüft (=

¹⁷Der PANAS wurde miterhoben, um für einen möglichen Effekt der Stimmung auf das DLC durch die NFCC Manipulation zu kontrollieren. Ein Regressionsmodell mit dem DLC als Kriterium und der (dummykodierten) NFCC Manipulation, dem positiven Affekt und dem negativen Effekt als Prädiktoren wird nicht signifikant und soll daher hier nicht näher beschrieben werden.

¹⁸In den Studien wurde Lärm über einen lauten Nadeldrucker im Labor, auf dem ein Stooze des Versuchsleiters angeblich dringend etwas ausdrucken muss, produziert.

3. Situative Einflussfaktoren

Manipulation-Check). In der dritten Phase wurde erneut das DLC gemessen, der unangenehme Ton und damit die situationale Manipulation des NFCC war hier abgestellt, um explorativ mögliche Kontrasteffekte zwischen den Bedingungen zu untersuchen (siehe weiter unten). In der abschließenden vierten Phase wurden Persönlichkeitsvariablen erhoben. Im Folgenden werden die Aufgaben der einzelnen Phasen genauer beschrieben.

Probanden nahmen in Einzelsitzungen an der Studie teil. Sie wurden von der Versuchsleiterin in das Labor geführt, in der sich sichtbar für die Probanden in einer Raumecke ein Computer befand - im Folgenden Mathecomputer genannt -, der angeschaltet war und auf dem Zahlenkolonnen und graphische Abbildungen auftauchten.

Nachdem sich die Probanden an einen zweiten Rechner gesetzt hatten - im Folgenden Studiencomputer genannt -, an dem die Studie stattfand, wurden sie schriftlich über ihre Rechte als Teilnehmer an einer psychologischen Studie aufgeklärt. Sie erhielten für die Teilnahme Geld (ca. 6 Euro) oder für ihr Studium notwendige Teilnahmestunden gutgeschrieben. Die Durchführung aller Aufgaben fand am Computer statt.

Die Versuchsleiterin gab während der gesamten Studie vor, am Laptop, den sie auf ihren Beinen liegen hatte und dessen Bildschirm von den Probanden abgewandt war, Vorlesungsfolien für eine Prüfung zu studieren.

Die Versuchsleiterin hatte jedoch für die Probanden unbemerkt über eine Software¹⁹ den Bildschirm des Studiencomputers und den Bildschirm des Mathecomputers auf dem Monitor.

Dann sollten die Probanden die 51 Durchgänge der Messaufgabe der Pferdewette bearbeiten, die in Kapitel 1.3.2.1 (S. 60) und Anhang A.2 (S. 171) beschrieben ist.

In dem Moment, in dem sich der Proband auf der letzten Seite zur Instruktion der Pferdewette-Aufgabe befand, aktivierte die Versuchsleiterin über Ihren Laptop in der experimentellen Bedingung der Studie einen lauten 1900 Herz Sinuston, der auf dem Matherechner abgespielt wurde. Die Dezibelzahl konnte aufgrund fehlender Messinstrumente nicht genau bestimmt werden. Die Tonlautstärke war konstant für alle Probanden der Bedingung so eingestellt, dass der Ton in einem Vorabtest von der Versuchsleiterin und einer weiteren unabhängigen Person als unangenehm aber *nicht* schmerzhaft empfunden wurde. Der Ton hatte, damit er nicht zu unnatürlich klang, eine leichte zyklische Amplitudenschwankung.

¹⁹<http://www.teamviewer.com/de/>, Zugriff 20.09.2010

Das Verhalten der Versuchsleiterin war ab diesem Zeitpunkt geskriptet und lief somit für alle Probanden der Experimentalbedingung gleich ab: Die Versuchsleiterin stand auf, ging zum Monitor des Matherechners, tippte auf den Computer und sagte verärgert: „Schon wieder, der Bildschirm ist kaputt! Leider kann ich nichts machen und darf den Computer nicht ausschalten, weil eine Simulation läuft.“ Es handelte sich um einen Apple Rechner, bei dem die Tonausgabe im Monitor verbaut ist, so dass nicht evident ist, dass der Ton durch Ausschalten eines Lautsprechers deaktiviert werden kann.

Dann ging die Versuchsleiterin zu ihrem Platz zurück und forderte die Probanden auf, „einfach weiterzuarbeiten“, weil sie jetzt nichts dagegen machen könne. Falls es erneut zu einer Nachfrage der Probanden kam, war die Versuchsleiterin instruiert zu sagen: „Ich kann nichts machen, bitte mach weiter, so gut es eben geht.“

Falls sich die Probanden mehrfach beschwerten oder den Wunsch äußerten, die Studie abzubrechen, sollte die Studie abgebrochen und die Probanden zu dem Inhalt der Studie aufgeklärt werden. Ein Studienabbruch kam aber nicht vor.

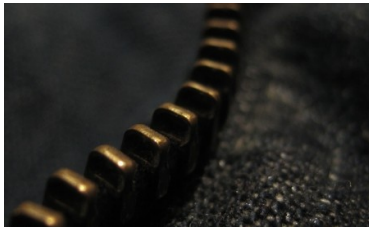
Dann bearbeiteten die Probanden die 51 Messdurchgänge der Pferdewette.

Dann folgte eine Manipulation-Check Aufgabe zum NFCC (Mayseless & Kruglanski, 1987). In der Aufgabe wurden per Zufallsreihenfolge zehn stark vergrößerte Ausschnitte von Alltagsobjekten wie einer Wasserflasche am Bildschirm präsentiert.²⁰ Die Probanden sollten zu jedem Bild erraten, welcher Alltagsgegenstand abgebildet ist. Dazu sollten sie alle Hypothesen, die ihnen dazu einfallen, aufschreiben. Nach jedem Bild bekamen die Probanden ein Feedback darüber, welcher Gegenstand abgebildet war; siehe Abbildung 3.1 (S. 116) für zwei Beispieldurchgänge der Aufgabe. Mayseless und Kruglanski (1987, Studie 3) konnten für diesen Aufgabentyp demonstrieren, dass Probanden mit einem erhöhten NFCC *seizing* Tendenzen zeigen: Sie generieren im Vergleich zu Probanden mit einem geringen NFCC weniger Hypothesen.

Der Ton wurde nach dem fünften Durchgang des Bilderrätsels von der Versuchsleiterin über ihren Laptop abgestellt, so dass die Probanden die letzten fünf Aufgaben der Bilderrätselaufgabe ohne Ton bearbeiteten. Dann bearbeiteten die Probanden noch einmal die 51 Messdurchgänge der Pferdewette. Die wiederholte Bearbeitung der Aufgaben erlaubt es,

²⁰weiterhin: Zange, Bleistiftanspitzer, Reißverschluss, Stift, Kopfhörer, Fahrradkatzenaugen, Banane, Vorhängeschloss, Spraydose

3. Situative Einflussfaktoren



(a) Beispieldurchgang 1



(b) Lösung 1



(c) Beispieldurchgang 2



(d) Lösung 2

Abbildung 3.1. Aufgabe Manipulation Check *Need For Cognitive Closure*. Es werden stark vergrößerte Alltagsobjekte präsentiert (a, c). Die Probanden sollen Hypothesen generieren, um welches Objekt es sich handelt. Danach werden die Lösungen zu den Durchgängen präsentiert (b, d).

explorativ mögliche Kontrasteffekte der Experimentalbedingung Lärm zu untersuchen. So ist es möglich, dass das DLC durch das Verlassen der aversiven Situation ansteigt und dann im Vergleich zum DLC in der Kontrollgruppe zum zweiten Messzeitpunkt signifikant höher liegt.

Dann wurden das *Bedürfnis nach kognitiver Geschlossenheit* (Schlink & Walther, 2007), der *regulatorische Fokus* (Werth & Förster, 2007) und die *Achievement Motives Scale* (Lang & Fries, 2006) am Computer erhoben; siehe Kapitel 4.2 (S. 135) für die Beschreibung der Skalen, Hypothesen des Zusammenhangs mit dem DLC und den Ergebnissen.

Im letzten Abschnitt der Studie wurden demographische Daten der Probanden erhoben.

Dann wurden die Probanden über den Inhalt der Studie und in der Bedingung Lärm über den Ursprung des Tons aufgeklärt und darüber befragt, ob sie während der Studie den Verdacht gehabt hatten, dass die Szene mit dem defekten Monitor gestellt gewesen war und zur Studie

gehört hatte.

Stichprobe Es nahmen nach Ausschluss von 16 Probanden, die angegeben, Verdacht geschöpft zu haben (siehe Methoden), 71 Probanden an der Studie teil. Davon waren 33 männlich und 38 weiblich. Das Alter lag zwischen 17 und 38 Jahren bei einem Mittelwert von 23.3 Jahren und einer Streuung von 4.2 Jahren.

3.2.2.2. Hypothesen

Probanden, die einem aversiven Ton ausgesetzt sind, haben ein niedrigeres DLC als Probanden der Kontrollgruppe.

3.2.2.3. Ergebnisse

Das mittlere DLC zu Messzeitpunkt 1 liegt bei .824 (SD = .117). Das mittlere DLC zu Messzeitpunkt 2 liegt bei .802 (SD = .129). Die Korrelation zwischen den beiden DLC Messungen liegt bei $r = .85$, $t(69) = 13.65$, $p < .05$.

Das mittlere DLC liegt zu Messzeitpunkt 1 in der Bedingung Lärm bei .835 (SD = .119). In der Kontrollgruppe liegt das mittlere DLC bei .812 (SD = .115). Der Unterschied ist mit einem $t(69) = .82$, $p = .79$ und Cohen's $d = 0.195$ nicht signifikant.²¹

Das mittlere DLC liegt zu Messzeitpunkt 2 in der Bedingung Lärm bei .814 (SD = .127). In der Kontrollgruppe liegt das mittlere DLC bei .789 (SD = .131). Der Unterschied ist mit einem $t(69) = .81$, $p = .41$ (ungerichtete Hypothese) und Cohen's $d = 0.193$ nicht signifikant.

Die mittlere Anzahl der generierten Lösung für die ersten fünf Durchgänge der Bilderrätsel-Aufgabe zum Manipulation Check des NFCC ergibt keine signifikanten Unterschiede zwischen der Bedingung Lärm und der Kontrollgruppe. Die mittlere Anzahl der generierten Lösungen liegt für die Bedingung Lärm bei 1.86 (SD = 1.06). Die mittlere Anzahl der generierten Lösungen liegt für die Kontrollgruppe bei 1.85 Lösungen (SD = 1.11). Der Unterschied ist mit einem $t(69) = .01$, $p = .50$ und Cohen's $d = 0.003$ nicht signifikant.

Die mittlere Anzahl der generierten Lösung für die letzten fünf Durchgänge der Bilderrätsel-Aufgabe zum Manipulation Check des NFCC, ergibt keine signifikanten Unterschiede zwischen der Bedingung Lärm

²¹Die Power liegt bei einer erwarteten mittleren Effektstärke von Cohen's $d = .5$, 71 Probanden und einem $\alpha = .05$ bei $1 - \beta = .67$.

3. Situative Einflussfaktoren

und der Kontrollgruppe. Die mittlere Anzahl der generierten Lösungen liegt für die Bedingung Lärm bei 1.92 Lösungen ($SD = 1.44$). Die mittlere Anzahl der generierten Lösungen liegt für die Kontrollgruppe bei 1.60 Lösungen ($SD = 1.01$). Der Unterschied ist mit einem $t(69) = 1.06$, $p = .29$ (ungerichtete Hypothese) und Cohen's $d = 0.25$ nicht signifikant.

3.2.3. Diskussion

Die Höhe des DLC lässt sich durch zwei verschiedene situationale Manipulationen des NFCC nicht verändern.

Es besteht in beiden Studien ein *statistisches* Problem: Die Power bei der Anzahl der erhobenen Probanden von $n = 40$ und $n = 71$, einer mittleren erwarteten Effektstärke und einem α -Fehler von .05 fällt mit einem Wert von .46 für die Manipulation des NFCC über die Aufgabenattraktivität und .67 für die Manipulation des NFCC über einen aversiven Ton gering aus, so dass die Wahrscheinlichkeit für einen β -Fehler in beiden Studien hoch ist.

In der Studie zur Manipulation des NFCC über die Aufgabeattraktivität führt die Aussicht Statistik vs. Comedy Videos zu schauen erwartungskonform zu geringerem positiven Affekt und höherem negativen Affekt. Es ist jedoch nicht klar bzw. es wurde nicht direkt gemessen, ob die Aufgabenattraktivität das interessierende Konstrukt, das ist das NFCC, beeinflusst, so dass es möglich ist, dass die Manipulation keine Wirkung hatte. In der Studie zur Manipulation des NFCC über einen aversiven Ton konnten NFCC Effekte auf die Manipulation-Check Aufgabe nicht nachgewiesen werden, so dass bezweifelt werden muss, dass die Manipulation ein erhöhtes NFCC hervorgerufen hat. Es lässt sich also kein Effekt des NFCC auf das DLC zeigen und auch nicht auf die von Kruglanski selbst verwendete Aufgabe des Manipulation-Checks (Kruglanski & Webster, 1991). Wenn die Manipulationen das situationale NFCC nicht verändert haben, dann kann folglich das NFCC auch keinen Einfluss auf das DLC haben.

Es besteht auch ein generelles konzeptuelles Problem: Laut Kruglanski und Webster (1991) treten bei hohem NFCC vor einer Entscheidung *seizing* Tendenzen und nach einer Entscheidung *freezing* Tendenzen auf. *Seizing* Tendenzen führen dazu, dass Entscheidungen basierend auf geringer Informationslage und schnell getroffen werden, um möglichst schnell den motivational aversiven Zustand der Ambiguität zu verlassen. *Freezing* Tendenzen führen dazu, dass eine einmal getroffene

Entscheidung nicht wieder revidiert wird, um nicht in den Zustand der Ambiguität zurückzufallen. Es ist jedoch fraglich, inwieweit diese gegenläufigen Prozesse streng sequentiell ablaufen und wenn das nicht der Fall ist, wie der Konflikt zwischen den beiden gegenläufigen Tendenzen aufgelöst wird. So ist es denkbar, dass Probandem mit hohem NFCC mehr Informationen suchen, um eine möglichst sichere Entscheidung zu treffen, und damit mehr Cues konsultieren und ein höheres DLC haben. Ein solcher Entscheider würde zwar kurzfristig den Zustand der Ambiguität, der vor der Entscheidung erlebt wird, langsamer verlassen können, aber auf lange Sicht weniger wahrscheinlich in den Zustand der Ambiguität zurückfallen, da die Entscheidung auf einer vollständigeren Informationslage getroffen wurde und damit die (zumindest subjektiv erlebte) Wahrscheinlichkeit erhöht sein sollte, dass die Entscheidung korrekt ist. Diesem Argument kann jedoch entgegengehalten werden, dass Probanden mit hohem NFCC trotz geringerer Informationssuche sogar tendenziell höhere Sicherheitsurteile für die Korrektheit ihrer Entscheidungen abgeben.²²

²²Die Autoren nennen diesen Befund "unfounded confidence paradox" (Kruglanski & Webster, 1991, S. 269).

4. Korrelation mit Persönlichkeitsvariablen

In den folgenden Studien wird untersucht, ob das DLC mit Persönlichkeitsvariablen zusammenhängt. Dazu werden in den ersten beiden Studien individuelle Antworttendenzen in einer Kategorisierungsaufgabe und in einer Rekognitionsaufgabe erhoben. Abschließend werden die Befunde der über die Studien hinweg erhobenen anderen Persönlichkeitsvariablen vorgestellt und diskutiert.

4.1. Signalentdeckung: Antworttendenzen

In den folgenden beiden Studien soll untersucht werden, ob das DLC mit dem Entscheidungskriterium der Signalentdeckung (Green & Swets, 1966; Macmillan & Creelman, 2005) zusammenhängt. Die Studie dienen dazu, die Konstruktvalidität des DLC zu bestimmen, indem überprüft wird, inwieweit verwandte Konstrukte miteinander korreliert sind. Im Folgenden werden daher nach einem kurzen Überblick zur Signalentdeckungstheorie die Zusammenhänge zwischen dem DLC und dem Entscheidungskriterium der Signalentdeckung dargestellt.

Gemäß der Signalentdeckungstheorie gibt es in einer Signalentdeckungsaufgabe einen Sender, der das Signal bzw. die Information verschickt, und einen Empfänger, der das Signal detektiert. Ein Sender bzw. Empfänger kann dabei ein Mensch oder eine Maschine sein. Das Signal kann durch internes oder externes Rauschen überdeckt sein, ein *sensitiver* Signalentdecker kann zwischen Durchgängen, in denen nur Rauschen vorliegt, und Durchgängen, in denen neben dem Rauschen ein Signal vorhanden ist, trennen.

Ein Signal kann z. B. ein Ton sein. Ein guter menschlicher Signalentdecker trennt die Durchgänge, in denen nur Rauschen vorliegt - intern kann Rauschen durch den Blutkreislauf im Ohr und extern kann Rauschen durch Umweltgeräusche entstehen - von den Durchgängen, in denen neben dem Rauschen ein Ton vorliegt.

4. Korrelation mit Persönlichkeitsvariablen

Ein Signalentdecker kann im Tonbeispiel eine von zwei¹ Antworten geben, nämlich „Signal“ vs. „kein Signal“ vorhanden; gleichzeitig kann in der Umwelt einer von zwei Zuständen vorliegen, nämlich das Signal war tatsächlich vorhanden vs. es war nicht vorhanden. Daraus ergeben sich vier mögliche Fälle: (1) Das Signal war vorhanden, der Signalentdecker sagt „Signal vorhanden“ oder (2) das Signal war nicht vorhanden und der Signalentdecker sagt „Signal nicht vorhanden“; (3) Das Signal war nicht vorhanden, der Signalentdecker sagt „Signal vorhanden“ oder (4) das Signal war vorhanden und der Signalentdecker sagt „Signal nicht vorhanden“. (1) und (2) sind korrekte Antworten und werden *Hits* und *Correct Rejections* genannt; (3) und (4) sind fehlerhafte Antworten und werden *False Alarms* und *Misses* genannt.

Ein Signalentdecker, der perfekt Signal von Rauschen trennen kann und somit perfekt *sensitiv* für den realen Zustand ist, sollte über die Durchgänge einer Signalentdeckungs-Aufgabe in 100% der Fälle *Hits* und in 100% der Fälle *Correct Rejections* produzieren. Ein Signalentdecker, der „perfekt“ *unsensitiv* für den realen Zustand ist (d. h., die Rauschenverteilung und die Signal + Rauschen-Verteilung liegen übereinander), sollte somit über die Durchgänge einer Signalentdeckungs-Aufgabe prozentual gleich viele *Hits* wie *False Alarms* und gleich viele *Correct Rejections* wie *Misses* produzieren.²

Die Stärke des Signals löst dabei im Signalentdecker eine „Empfindung“ (Lukas, 2006, S. 733) - das ist die subjektive Wahrnehmung der Signalstärke - aus. Dabei sollte die Empfindungsstärke bei Durchgängen, in denen ein Signal vorliegt, im Mittel höher liegen als bei Durchgängen, bei denen kein Signal vorliegt. Unter der Annahme, dass die Empfindungen für beide möglichen Zustände - Signal liegt vor und Signal liegt nicht vor, normalverteilt ist, lässt sich die Sensitivität oder auch „Trennschärfe“ (Lukas, 2006, S. 735) - d' - eines Signalentdeckers anhand der *Hits* - H - und der *False Alarms* - F - nach folgender Formel berechnen (Stanislaw & Todorov, 1999, Formel 1, S. 142):

$$d' = \Phi^{-1}(H) - \Phi^{-1}(F) \quad (4.1)$$

¹Im Tonbeispiel gibt es zwei mögliche Zustände, die hier vorgestellten Maße der Signalentdeckung können jedoch auch auf Situationen mit mehr als zwei Zuständen erweitert werden (Macmillan & Creelman, 2005).

²Prinzipiell ist ein Ergebnis von 100% *Misses* und 100% *False Alarms* möglich, das bedeutet aber, dass der Signalentdecker fälschlicherweise die Evidenz für ein Signal als Evidenz gegen das Signal und damit fälschlicherweise als Evidenz für Rauschen interpretiert.

Die Transformationsfunktion Φ^{-1} übersetzt die relativen Häufigkeiten bzw. Wahrscheinlichkeiten der *Hits* bzw. der *False Alarms* in die korrespondierenden z -Werte der Standardnormalverteilung. Je sensibler ein Proband ist, desto größer ist die Differenz zwischen dem Mittelwert der Normalverteilung der Empfindungsstärke der Durchgänge, in denen ein Signal vorliegt, von dem Mittelwert der Normalverteilung der Empfindungsstärke der Durchgänge, in denen kein Signal vorliegt.

Ob ein Signalentdecker sich für den Zustand „Signal“ vs. „kein Signal“ entscheidet, hängt jedoch davon ab, wie individuell stark die Empfindungsstärke ausgeprägt sein muss, bevor sich ein Signalentdecker für den Zustand „Signal“ entscheidet. Die gleiche Empfindungsstärke kann also für Signalentdecker, je nachdem ob die Empfindungsstärke ein individuelles Entscheidungskriterium bzw. eine individuelle Entscheidungsschwelle übertrifft, zu unterschiedlichen Antworten führen.

Das Entscheidungskriterium - $\ln(\beta)$ - lässt sich über folgende Formel berechnen (Stanislaw & Todorov, 1999, Formel 6, S. 142):

$$\ln(\beta) = \frac{[\Phi^{-1}(F)]^2 - [\Phi^{-1}(H)]^2}{2} \quad (4.2)$$

Ein Signalentdecker, bei dem die Empfindungsstärke sehr hoch sein muss, bevor er sich für den Zustand „Signal“ entscheidet, hat ein *kon-servatives* Entscheidungskriterium (Lukas, 2006), was sich in einem positiven Wert für $\ln(\beta)$ widerspiegelt.

Ein Signalentdecker, bei dem eine geringe Empfindungsstärke ausreicht, damit er sich für den Zustand „Signal“ entscheidet, hat ein *liberales* Entscheidungskriterium (Lukas, 2006), was sich in einem negativen Wert für $\ln(\beta)$ widerspiegelt.

Wenn ein Signalentdecker genauso viele *Correct Rejections* wie *Hits* produziert, hat er mit einem $\ln(\beta)$ von 0 ein neutrales Entscheidungskriterium.

Das Brunswiksche Linsenmodell (S. 22 in Kapitel 1.1) lässt sich mit der Signalentdeckung konzeptuell verknüpfen: Die Ausprägung der distalen Variablen ist der reale Zustand, d. h., ob ein Signal vorliegt oder nicht vorliegt. Die zentrale Antwort eines Entscheiders ist das Urteil über den realen Zustand, d. h., ob ein Signalentdecker meint, dass ein Signal vorliegt oder nicht. Die subjektive Empfindung der Signalstärke basiert auf den Ausprägungen von Cues³ und der Cue-Nutzung eines

³siehe z. B. Zuckerman, Koestner und Driver (1981, Table 1, S. 111) und DePaulo

4. Korrelation mit Persönlichkeitsvariablen

Entscheidungers. Wie sensitiv ein Signalentdecker für die realen Zustände ist, spiegelt sich in der Passung zwischen den ökologischen Validitäten und der Cue-Nutzung wider.

Über die Signalentdeckungstheorie lässt sich die Sensitivität eines Entscheiders und das angelegte Entscheidungskriterium berechnen; über das Brunswiksche Linsenmodell als statistisches Tool lässt sich anhand der ökologischen Validitäten der Cues und der Cue-Nutzung analysieren, welche Cues mit einer Entscheidung zusammenhängen - d. h. also, welche Cues mit der Empfindungsstärke korrelieren - und weshalb ein Signalentdecker eine hohe oder geringe Sensitivität hat. Die Sensitivität sollte dabei mit der Passung zwischen der Cue-Nutzung und der ökologischen Validität der Cues korrelieren. Zu beachten ist hier jedoch, dass die kausale Richtung eines Cues nur experimentell erfasst werden kann, da ein Cue *nur* als Beiprodukt eines anderen kausal wirksamen Cues mit dem Urteil eines Probanden korrelieren kann und somit nicht ursächlich für die Empfindungsstärke sein muss.

Das Signalentdeckungsmodell ist also wie das DLC Modell ein Schwellenmodell einer Entscheidung.⁴ Eine Entscheidung basiert im DLC Modell auf probabilistischen Cues, deren Ausprägung und Nutzung in der Empfindungsstärke im Signalentdeckungs-Modell repräsentiert ist. Wie stark und damit wie valide eine Empfindung sein muss, damit ein Entscheider sich für den Zustand „Signal liegt vor“ entscheidet, hängt von einem individuell messbaren Entscheidungskriterium ab.

Das Entscheidungskriterium der Signalentdeckungstheorie und das DLC sind also konzeptuell verwandt, ihr Zusammenhang soll in den folgenden beiden Studien empirisch untersucht werden.

4.1.1. Zuordnungsaufgabe: Objekte klassifizieren

In der folgenden Studie wird der Zusammenhang des DLC mit dem Entscheidungskriterium der Signalentdeckungstheorie in einer Kategorisierungsaufgabe untersucht.

et al. (2003) für Cues in der Detektion einer Lüge (= Signal) - wie z. B. Augenblinzeln, Adaptoren, Schulterzucken, Antwortlatenz, Stimmhöhe, Selbstreferenz - und Fiedler und Walka (1993) für die Analyse der Detektion einer Lüge im Brunswikschen Linsenmodell

⁴siehe Swets, Dawes und Monahan (2000) für die Relevanz von Signalentdeckung in der Praxis und der Signalentdeckungstheorie als generelles Tool zur Analyse von Entscheidungen

4.1.1.1. Methoden

Die Studie bestand aus vier Phasen. In der ersten Phase wurde das DLC erhoben. Dann bearbeiteten die Probanden eine Kategorisierungsaufgabe, in der über das Signalentdeckungsmodell das (individuelle) Entscheidungskriterium der Probanden ermittelt wurde. Dann bearbeiteten die Probanden erneut die Bilderrätselaufgabe (siehe Studie 2 auf S. 69), die zur Replikation des Korrelationsbefundes⁵ aus der vorherigen Studie aufgenommen wurde. Abschließend wurden Persönlichkeitsvariablen erhoben. Im Folgenden werden die Aufgaben der einzelnen Phasen genauer beschrieben.

Die Probanden wurden schriftlich über ihre Rechte als Teilnehmer an einer psychologischen Studie aufgeklärt. Die Durchführung aller Aufgaben fand am Computer statt. Die Probanden erhielten für ihre Teilnahme für ihr Studium notwendige Versuchspersonenstunden gutgeschrieben und die drei Probanden mit der höchsten Punktzahl (s. u.) zusätzlich einen Kinogutschein.

Dann bearbeiteten die Probanden die 71 Durchgänge der Pferdewette, die in Kapitel 1.3.2.1 (S. 60) und Anhang A.2 (S. 171) beschrieben ist.

Dann folgte die Kategorisierungsaufgabe, die aus einer Studie von Meyer, Feinshreiber und Parmet (2003) übernommen wurde. In der Aufgabe sollten die Probanden als Geschäftsführer einer Handyfirma entscheiden, ob ein Handy aus ihrer Produktion defekt ist und entsorgt werden muss oder funktionsfähig ist und in den Verkauf gehen kann.

Die Handys sind als 5×5 Matrix bestehend aus weißen und schwarzen Elementen repräsentiert. Ein Handy ist defekt, wenn vier schwarze Elemente ein Quadrat bilden. In Abbildung 4.2 (S. 127) sind bei Handy A alle Elemente weiß, das Handy ist funktionsfähig und sollte in den Verkauf gehen. Bei Handy B sind vier Elemente defekt *und* bilden ein Quadrat. Das Handy ist defekt und sollte entsorgt werden.

Damit es sich um eine Entscheidung über den Zustand eines Handys um eine Entscheidung unter Unsicherheit handelt, kann ein schwarzes Element zwei Dinge bedeuten: a) Die Qualitätskontrolle der Handyfirma hat das Element überprüft und es ist fehlerhaft oder b) das Element wurde aus Zeitdruck nicht überprüft, es also nicht sicher ist, ob das Element fehlerfrei ist; siehe Anhang B.5 (S. 179) für die verwendete Instruktion.

⁵Die Anzahl der geöffneten Maskenelemente war zumindest mit dem DLC des Kriminalfalls signifikant korreliert.

4. Korrelation mit Persönlichkeitsvariablen

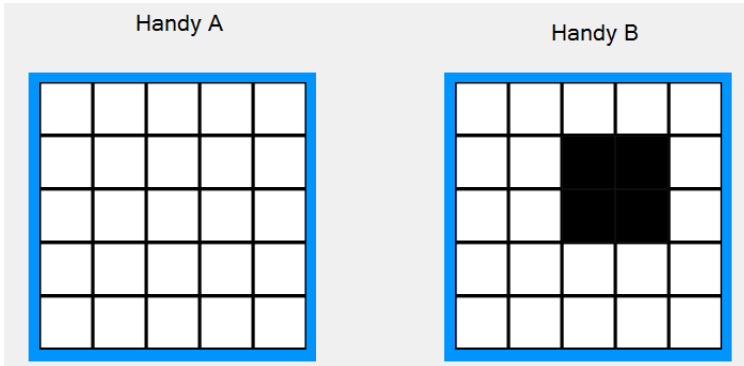


Abbildung 4.1. Gezeigt werden zwei Handys, bei denen alle Elemente kontrolliert wurden. Handy A ist fehlerfrei, Handy B hat vier schwarze fehlerhafte Elemente, die ein Quadrat bilden, und ist somit defekt.

In Abbildung 4.2 (S. 127) sind zwei Handys abgebildet, wie sie die Probanden in der Aufgabe sehen.

Handy C ist eindeutig fehlerfrei. Selbst wenn die schwarzen Elemente fehlerhaft sind, bilden die schwarzen Elemente kein Quadrat. Der Zustand von Handy D ist nicht eindeutig. Es gibt vier schwarze Elemente die ein Quadrat bilden, wenn die Elemente tatsächlich fehlerhaft sind, ist das Handy defekt, wenn aber nur eines der Elemente des schwarzen Quadrates fehlerfrei ist und nur nicht kontrolliert wurde, ist das Handy funktionsfähig. Über die Anzahl der weißen Felder kann die Schwierigkeit der Aufgabe manipuliert werden.

Es wurden insgesamt 96 Handys präsentiert. Von den 96 Handys waren 48 Handys funktionsfähig und 48 Handys defekt. Für die 48 defekten Handys wurde das schwarze Quadrat per Zufall auf der 5×5 Matrix positioniert. Die Schwierigkeit wurde in drei Stufen variiert, indem die funktionsfähigen und die defekten Handys per Zufallsposition mit unterschiedlich vielen schwarzen Elementen aufgefüllt wurden, so dass das Aufgabenset aus 2 (defekt vs. funktionsfähig) * 3 (10, 14 oder 18 weiße Felder sind zu sehen) * 16 (Handys pro Kategorie) Handys besteht.

In jedem Durchgang sollten die Probanden entscheiden, ob das Handy „defekt“ oder „in Ordnung“ ist; siehe Abbildung 4.3 (S. 128) für die Oberfläche der Aufgabe, wie sie die Probanden gesehen haben.

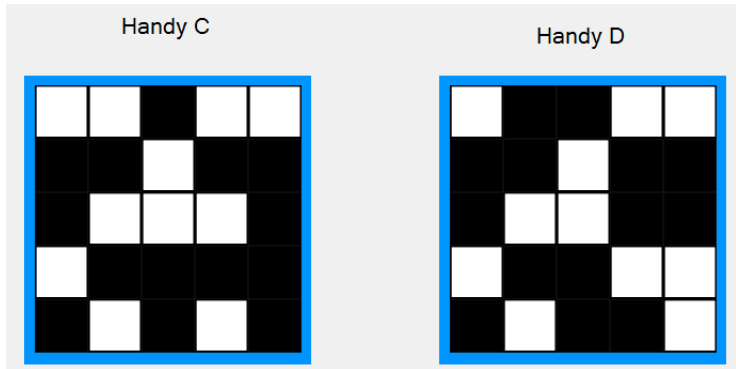


Abbildung 4.2. Gezeigt werden zwei Handys, bei denen *nicht* alle Elemente kontrolliert wurden. Ein schwarzes Element kann bedeuten, dass das Element fehlerhaft ist oder dass das Feld nicht kontrolliert wurde. Handy C ist funktionsfähig, Handy D kann defekt sein (siehe Text).

Den Probanden wurde nicht die Anzahl der Durchgänge mitgeteilt, jedoch die Fehlerrate von 50%. Nach jeder Entscheidung gab es ein Feedback. Für eine korrekte Entscheidung - das sind *Hits* und *Correct Rejections* - wurde den Probanden 100 Euro Spielgeld auf ihrem Konto gutgeschrieben. Für eine inkorrekte Entscheidung - das sind *Misses* und *False Alarms* - wurde den Probanden 100 Euro Spielgeld von ihrem Konto abgezogen.

In Anlehnung an Meyer et al. (2003) ist ein *Hit* als Detektion eines defekten Handys, eine *Correct Rejection* als korrekte Zurückweisung eines funktionsfähigen Handys als defekt, ein *Miss* als verpasste Detektion eines defekten Handys und ein *False Alarm* als inkorrekte Klassifikation eines funktionsfähigen Handys als defekt definiert. Das bedeutet also, dass ein *negatives* Entscheidungskriterium $\ln(\beta)$ für ein „risikoaversives“ Entscheidungskriterium steht (Meyer et al., 2003): Wenn ein *Hit* die Detektion eines defekten Handys und eine *Correct Rejection* die Zurückweisung eines funktionsfähigen Handys ist, dann tendiert ein Proband mit einem negativen $\ln(\beta)$ dazu, bei geringer Signalstärke „defekt“ zu antworten. Ein solcher Proband minimiert also das Risiko, ein defektes Handy fälschlicherweise als funktionsfähig zu deklarieren.⁶

⁶Meyer et al. (2003, S. 2104) „Higher settings of the criterion indicate a greater

4. Korrelation mit Persönlichkeitsvariablen

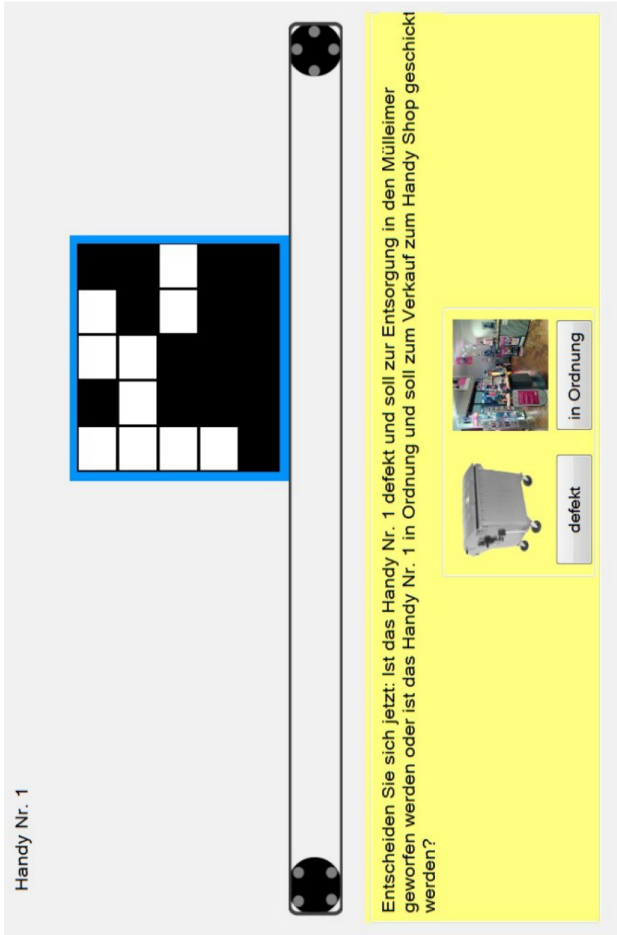


Abbildung 4.3. Die Oberfläche der Kategorisierungsaufgabe. Ein Proband muss entscheiden, ob das Handy auf dem Fließband fehlerhaft ist und entsorgt werden muss oder fehlerfrei ist und verkauft werden kann.

Dann bearbeiteten die Probanden in 40 Durchgängen eine Bilderrätsel Aufgabe, in der Gegenstände mit einer Maske bestehend aus 100 Elementen verdeckt ist, ein Element alle 400 Millisekunden aufgedeckt

tendency to take risks and to declare that an item is intact.“

wird und die Probanden so schnell wie möglich erraten sollen, welches Objekt unter der Maske abgebildet ist; siehe S. 72 in Kapitel 2.1 für Details.

Dann wurden das *Bedürfnis nach kognitiver Geschlossenheit* (Schlink & Walther, 2007), der *regulatorische Fokus* (Werth & Förster, 2007) und die *Achievement Motives Scale* (Lang & Fries, 2006) am Computer erhoben; siehe Kapitel 4.2 (S. 135) für die Beschreibung der Skalen, Hypothesen des Zusammenhangs mit dem DLC und den Ergebnissen.

In der letzten Phase der Studie wurden demographische Daten der Probanden erhoben.

Stichprobe Es nahmen 48 Probanden an der Studie teil. Davon waren 20 männlich und 28 weiblich. Das Alter lag zwischen 13 und 75 Jahren bei einem Mittelwert von 31.93 Jahren⁷ und einer Streuung von 15.30 Jahren.

4.1.1.2. Hypothesen

Es gibt einen negativen Zusammenhang zwischen dem DLC und dem Entscheidungskriterium $\ln(\beta)$ der Signalentdeckungstheorie.

Probanden mit einem höheren DLC benötigen mehr Evidenz im Bilderrätsel, bevor sie einen Tipp abgeben. Es wird eine negative Korrelation zwischen dem DLC der Pferdewette und der Punktezahl des Bilderrätsels erwartet.

4.1.1.3. Ergebnisse

Das mittlere DLC der Pferdewette liegt bei .784 bei einer Streuung von .126. Die Modelltestung gegen den Zufall⁸ ergibt, dass 38 von 48 (79%) Probanden als Anwender des DLC als Stopppregel der Informationssuche in den 20 Testdurchgängen der Pferdewette klassifiziert werden können. 4 Probanden können als Anwender einer ORDM-, 2 Probanden als Anwender einer MRDM- und 1 Proband als Anwender einer Zufalls-Stopppregel klassifiziert werden. 3 Probanden lassen sich nicht eindeutig nur einem der Modelle zuordnen.

Die Sensitivität d' liegt im Mittel bei 1.53 bei einer Streuung von 0.53. Das Entscheidungskriterium $\ln(\beta)$ liegt im Mittel bei -0.48 bei

⁷n = 47, da ein Proband eine fehlerhafte Eingabe („2“) beim Alter gemacht hat.

⁸siehe S. 64 in Kapitel 1.3.2.2 für Details zur Testung gegen den Zufall

4. Korrelation mit Persönlichkeitsvariablen

einer Streuung von 1.01. Die Hypothese, dass das DLC mit dem Entscheidungskriterium $\ln(\beta)$ negativ korreliert ist, kann nicht bestätigt werden. Probanden mit einem hohen DLC haben kein signifikant niedrigeres Entscheidungskriterium $\ln(\beta)$, d. h., Probanden mit einem hohen DLC sind in ihrer Entscheidung nicht signifikant konservativer. Die Korrelation ist mit Pearson's $r = .03$, $t(46) = 0.25$, $p = .60$ nicht⁹ signifikant.¹⁰ Es zeigt sich jedoch, dass zwei Probanden mit einer vierfachen Standardabweichung unter dem mittleren Entscheidungskriterium liegen. Reduziert man den Datensatz um die beiden Messwerte, ergibt die Korrelation zwischen dem DLC und dem Entscheidungskriterium $\ln(\beta)$ einen nicht signifikanten Zusammenhang mit Pearson's $r = .10$, $t(44) = -0.68$, $p = .75$, der in die hypotheseninkonforme Richtung zeigt.

Die Hypothese über den negativen Zusammenhang zwischen dem DLC und der Punktzahl der Bilderrätsel-Aufgabe kann nicht bestätigt werden. Es zeigt sich zwar deskriptiv, dass die Probanden mit einem hohen DLC dazu tendieren, mehr Kästchen in der Bilderrätsel-Aufgabe aufzudecken, bevor sie einen Tipp über das Objekt unter der Maske abgeben, der Zusammenhang ist jedoch nicht signifikant von 0 verschieden. Cronbach's α der Punktezahlen der 40 Durchgänge des Bilderrätsels liegt bei .87. Die hohe interne Konsistenz der Aufgabe spricht dafür, dass die 40 Durchgänge das gleiche Konstrukt erfassen, so dass die Rechnung mit der gemittelten Punktzahlen gerechtfertigt ist. Die Korrelation zwischen dem DLC und der mittleren Punktzahl des Bilderrätsels liegt bei Pearson's $r = -.19$, $t(46) = -1.35$, $p = .09$.

Es lässt sich also kein Zusammenhang zwischen dem Entscheidungskriterium in einer Kategorisierungsaufgabe und dem DLC zeigen. Die Befunde werden nach der nächsten Studie zusammenfassend diskutiert.

4.1.2. Rekognition: Gesichter wiedererkennen

In der folgenden Studie wird der Zusammenhang des DLC mit dem Entscheidungskriterium der Signalentdeckungstheorie in einer Rekognitionsaufgabe untersucht. Gemessen wird wiederum das Entscheidungskriterium und mit dem DLC verglichen.

⁹Die Power liegt bei einer erwarteten mittleren Effektstärke von Pearson's $|r| = .3$ und einer Stichprobengröße von 52 Probanden und einem $\alpha = .05$ bei $1 - \beta = .69$

¹⁰Zur Komplementierung der Befunde: Die Korrelation zwischen dem DLC und der Sensitivität d' liegt bei Pearson's $r = -.05$, $t(46) = -0.37$, $p = .70$.

4.1.2.1. Methoden

Die Studie bestand aus drei Phasen. Die erste Phase bestand aus der Lernphase der Rekognitionsaufgabe. Die zweite Phase bestand aus der Messung des DLC. Die dritte Phase bestand aus der Testphase der Rekognitionsaufgabe. Anhand der Rekognitionsaufgabe lässt sich über das Signalentdeckungsmodell das (individuelle) Entscheidungskriterium der Probanden berechnen und dann mit dem DLC vergleichen. Im Folgenden werden die Aufgaben der einzelnen Phasen genauer beschrieben.

Die Probanden wurden schriftlich über ihre Rechte als Teilnehmer an einer psychologischen Studie aufgeklärt. Die Durchführung aller Aufgaben fand am Computer statt. Die Probanden erhielten für ihre Teilnahme für ihr Studium notwendige Versuchspersonenstunden gutgeschrieben.

In der ersten Phase der Studie wurden den Probanden 90 Frontalphotographien von kaukasischen Gesichtern mit einem emotionsneutralem Gesichtsausdruck vor einem neutralen weißen Hintergrund in einer Bildgröße von 641×480 Pixeln präsentiert.¹¹ Die Anzahl der abgebildeten Frauen und Männern und der Altersgruppen 18-30 Jahre, 31-50 Jahre, > 50 Jahre waren über das Set gleichverteilt. Die Photographien wurden per Zufallsreihenfolge in einem zeitlichen Abstand von 1.5 Sekunden eingespielt. Die Probanden wurden instruiert, sich die Gesichter gut einzuprägen, und sie wurden über die später in der Studie folgende Testphase der Rekognitionsaufgabe informiert.

Nach der Lernphase der Signalentdeckungsaufgabe bearbeiteten die Probanden die 71 Durchgänge der Pferdewette, die in Kapitel 1.3.2.1 (S. 60) und Anhang A.2 (S. 171) beschrieben ist. In Abweichung zu den vorigen Studien wurde aus Gründen der Flexibilität eine PHP/HTML Version der Pferdewette programmiert (siehe Abbildung C.1 auf S. 193 in Anhang C.5), die aber in dem Design der Oberfläche identisch zur Visual Basic Version der Pferdewette ist (siehe Abbildung 1.4 auf S. 61 in Kapitel 1.3.2.1).

Falls die Probanden weniger als 25 Minuten für die Bearbeitung der Pferdewette benötigten, sollten sie in der restlichen Zeit Sudokus lösen.

Dann folgte die Testphase der Signalentdeckungsaufgabe. Dazu wurden zu den 90 Photographien aus der Lernphase 90 Photographien von Gesichtern (= Distraktoren)¹² hinzugefügt und alle Gesichter per Zu-

¹¹Die Photos wurden im Rahmen einer Studie in der Abteilung Sozialpsychologie, Psychologisches Institut, Universität Heidelberg erstellt.

¹²Das Set der Zielstimuli und Distraktoren wurde nicht ausbalanciert, da interindi-

4. Korrelation mit Persönlichkeitsvariablen

fallsreihenfolge eingespielt. Die Merkmale der Photographien (s. o.) der Lern- und der Testphase waren identisch. Die Aufgabe der Probanden war es zu entscheiden, ob das Gesicht schon einmal in der Lernphase präsentiert worden waren. Es gab kein Feedback über die Akkuratheit der Entscheidung. Ein *Hit* ist als Rekognition eines Gesichtes aus der Lernphase, eine *Correct Rejection* als eine korrekte Klassifikation eines Distraktors, ein *Miss* als eine fälschliche Klassifikation eines Gesichtes aus der Lernphase als Distraktor und ein *False Alarm* als inkorrekte Klassifikation eines Distraktors als Gesicht aus der Lernphase definiert. Wenn ein *Hit* die Detektion eines in der Lernphase gesehenen Gesichtes und eine *Correct Rejection* die Zurückweisung eines Distraktor Reizes (d. h., eines zuvor nicht gesehenen Gesichtes) ist, dann tendiert ein Proband mit einem hohen $\ln(\beta)$ dazu, erst bei einer hohen Signalstärke zu antworten, das Gesicht zuvor in der Lernphase gesehen zu haben. Ein solcher Proband minimiert also das Risiko, ein Gesicht, das neu in der Testphase eingespielt wird, als schon in der Lernphase gesehenes Gesicht zu deklarieren.

In der letzten Phase der Studie wurden demographische Daten der Probanden erhoben.

Stichprobe Es nahmen 52 Probanden an der Studie teil. Davon waren 17 männlich und 34 weiblich, ein Proband hat keine Angabe zum Geschlecht gemacht. Das Alter lag zwischen 14 und 67 Jahren bei einem Mittelwert von 29.6 Jahren und einer Streuung von 13.5 Jahren.

4.1.2.2. Hypothesen

Es gibt einen positiven Zusammenhang zwischen dem DLC und dem Entscheidungskriterium $\ln(\beta)$ der Signalentdeckungstheorie.

4.1.2.3. Ergebnisse

Das mittlere DLC der Pferdewette liegt bei .799 bei einer Streuung von .127. Die Modelltestung gegen den Zufall¹³ ergibt, dass 40 von 52 (76.92%) Probanden als Anwender des DLC als Stopregel der Informationssuche in den 20 Testdurchgängen der Pferdewette klassifiziert werden können. 0 Probanden können als Anwender einer ORDM-, 9

viduelle Unterschiede nicht durch unterschiedliche Photosets als mögliche Quelle für Fehlervarianz beeinflusst werden sollten.

¹³siehe S. 64 in Kapitel 1.3.2.2 für Details zur Testung gegen den Zufall

Probanden als Anwender einer MRDM- und 2 Probanden als Anwender einer Zufalls-Stoppregel klassifiziert werden. 1 Proband lässt sich nicht eindeutig nur einem der Modelle zuordnen.

Die Sensitivität d' liegt im Mittel bei 0.63 bei einer Streuung von 0.29. Das Entscheidungskriterium $\ln(\beta)$ liegt im Mittel bei 0.08 bei einer Streuung von 0.29. Die Hypothese, dass das DLC mit dem Entscheidungskriterium $\ln(\beta)$ positiv korreliert ist, kann nicht bestätigt werden. Probanden mit einem hohen DLC haben kein signifikant erhöhtes Entscheidungskriterium $\ln(\beta)$, d. h., Probanden mit einem hohen DLC sind in ihrer Entscheidung nicht signifikant konservativer. Die Korrelation ist mit Pearson's $r = -.15$, $t(50) = -1.10$, $p = .86$ nicht signifikant¹⁴ und in der Richtung hypotheseninkonform.¹⁵

4.1.3. Diskussion

Die Studien zeigen für zwei verschiedene Inhaltsklassen - Klassifikation und Rekognition - für eine *forced choice* Signalentdeckungsaufgabe mit Ja/Nein Antwortformat (Macmillan & Creelman, 2005, S. 3) keinen signifikanten Zusammenhang zwischen dem DLC und dem Entscheidungskriterium $\ln(\beta)$.

Es besteht in beiden Studien ein *statistisches* Problem: Die Power bei der Anzahl der erhobenen Probanden von $n = 52$ und $n = 39$, einer mittleren erwarteten Effektstärke und einem *alpha*-Fehler von .05 fällt mit einem Wert von .69 für die Kategorisierungsaufgabe und .72 für die Rekognitionsaufgabe gering aus, so dass die Wahrscheinlichkeit für einen β -Fehler in beiden Studien hoch ist.

Weiterhin ist nicht klar, inwieweit es sich bei dem Entscheidungskriterium in der Signalentdeckung um ein zeitstabiles Maß handelt. Das Entscheidungskriterium kann aus statistischen Gründen instabil sein, d. h., dass das Messinstrument (die Aufgabe) fehlerbehaftet ist, oder aus konzeptuellen Gründen instabil sein, d. h., dass es sich bei dem Entscheidungskriterium um eine zeitinstabile Eigenschaft und damit nicht um eine zeitstabile Persönlichkeitseigenschaft handelt. Wenn das Entscheidungskriterium nicht reliabel ist, kann es auch nicht mit einem anderen Maß wie dem DLC hoch korrelieren.¹⁶ Zwar gibt es Studien ,

¹⁴Die Power liegt bei einer erwarteten mittleren Effektstärke von $|\text{Pearson's } r| = .3$ und einer Stichprobengröße von 52 Probanden und einem $\alpha = .05$ bei $1 - \beta = .72$

¹⁵Zur Komplementierung der Befunde: Die Korrelation zwischen dem DLC und der Sensitivität d' liegt bei Pearson's $r = .24$, $t(50) = 1.76$, $p = .08$.

¹⁶siehe dazu die doppelte Minderungskorrektur (Amelang & Zielinski, 2002, Formel 2.22, S. 43)

4. Korrelation mit Persönlichkeitsvariablen

die zeigen, dass sich das Entscheidungskriterium durch z. B. die Payoff Struktur beeinflussen lässt (siehe z. B. Bröder und Schütz (2009) für einen Überblick), dem Autor dieser Arbeit sind jedoch keine Studien bekannt, die die zeitliche Stabilität des Entscheidungskriteriums untersuchen. Es ist also notwendig, die Stabilität des Entscheidungskriteriums $\ln(\beta)$ zu erfassen bzw. deren Vorhandensein abzusichern, bevor die konzeptuellen Gründe einer fehlenden Interkorrelation zwischen DLC und $\ln(\beta)$ diskutiert werden können.¹⁷

Vergleicht man die DLC Aufgabe mit den Signalentdeckungsaufgaben, gibt es Unterschiede im Aufgabentyp: In der DLC Aufgabe müssen die Probanden entscheiden, bei welcher von *mehreren* Optionen - d. h. welches Pferd - ein Signal vorliegt - das Gewinnerpferd ist. In den beiden verwendeten Signalentdeckungsaufgaben muss dagegen über das Vorhandensein eines Signals bzw. über die Abwesenheit eines Signals bei *einem* Objekt/Stimulus entschieden werden.

In der DLC Aufgabe wird also die relative Evidenz für eine Option und in der Signalentdeckungsaufgabe die absolute Evidenz für eine Option präsentiert. Es ist möglich, dass sich die Schwelle für eine relative von einer Schwelle für eine absolute Evidenz konzeptuell unterscheidet.

Die DLC Aufgabe lässt sich den verwendeten Signalentdeckungsaufgaben angleichen, indem nur ein Pferd pro Durchgang präsentiert wird und die Probanden basierend auf einer sequentiellen Konsultation von Cues, die mit einer bestimmten Validität die korrekte Antwort „Siegerpferd“ anzeigen, entscheiden sollen, ob das Pferd ein Gewinnerpferd ist. In diesem Fall bedeutet ein hohes DLC, dass die Validität eines Cues für die Entscheidung „Gewinnerpferd“ bzw. „Signal liegt vor“ hoch sein muss. Im jetzigen Aufgabentyp des DLC bedeutet ein hohes DLC dagegen, dass ein Cue, dessen Validität unterhalb der DLC Schwelle liegt, nicht ausreicht, um sich im Vergleich zu den anderen drei Optionen für die vom Cue angezeigte Option zu entscheiden und weitere Evidenz/Cues konsultiert werden muss/müssen.

Es stellt sich weiterhin die Frage, ob das DLC und das Entscheidungskriterium $\ln(\beta)$ gleichermaßen durch motivationale Anteile bestimmt wird. Laut dem theoretischen Hintergrund des DLC (siehe Theorieteil der Arbeit), bildet das DLC eine emotional-motivationale Komponente im Entscheidungsprozess ab: Das DLC ist das Anspruchsniveau an gewünschter Urteilssicherheit, das durch die Information erreicht werden muss, bevor die Suche nach weiteren Informationen abgebrochen

¹⁷d. h. also auch, dass die Zurodnung der beiden Studien in das Unterkapitel Persönlichkeitsvariablen in dieser Arbeit zum jetzigen Zeitpunkt gewagt ist

wird und eine Entscheidung getroffen wird (Hausmann & Läge, 2008). Es ist bei dem Entscheidungskriterium $ln(\beta)$ unklar, zu welchem Verhältnis motivationale und rein kognitiv-strategische Komponenten wirken. Für den hypothetischen Extremfall, dass dem DLC nur emotional-motivationale Komponenten und dem Entscheidungskriterium $ln(\beta)$ nur kognitiv-strategische Komponenten zugrunde liegen, kann die Korrelation zwischen DLC und dem Entscheidungskriterium $ln(\beta)$ niedrig ausfallen, wenn emotional-motivationale und kognitiv-strategische Komponenten unabhängig voneinander sind. Diese Idee ist jedoch hochspekulativ, da es keine Hinweise für die emotional-motivationalen und strategisch-kognitiven Komponenten gibt bzw. es keine Möglichkeit in den Studien gibt, beide Anteile für eine bestimmte Ausprägung eines DLCs und eines Entscheidungskriteriums $ln(\beta)$ in den Aufgaben getrennt zu messen.¹⁸ Weiterhin wurde auch in beiden Studien absichtlich die Payoff-Struktur (in der Kategorisierungsaufgabe) und die gleiche Basisrate für jeden Zustand (Signal da/nicht da) in den Studien für die unterschiedlichen möglichen korrekten und inkorrekten Entscheidung in der Signalentdeckungsaufgabe gleich hoch gesetzt. Dadurch wird nicht per se durch die Situation das Entscheidungskriterium in eine bestimmte Richtung gedrängt, wie z. B. aus strategischen Gründen eine bestimmte Antwort („Signal“/„kein Signal“) zu bevorzugen. Die Persönlichkeitskomponenten wie z. B. motivationale Komponenten sollten dann eine höhere Chance haben, wirksam zu werden.

4.2. Andere Persönlichkeitsvariablen

Über die berichteten Studien hinweg wurden Persönlichkeitsmaße erhoben. Die Befunde sollen hier zusammengefasst und diskutiert werden.

Die Skala zur Messung des *Bedürfnisses nach kognitiver Geschlossenheit* - der 16-NCCS (Schlink & Walther, 2007) - misst das chronische *Need For Cognitive Closure* einer Person; siehe Kapitel 3.2 (S. 108). Somit wurden neben den situativen Manipulationen des NFCC dieser Studie, auch der Zusammenhang zwischen einem stabilen NFCC und dem DLC untersucht.

Probanden mit einem hohen Wert auf der Skala tendieren dazu, schnelle und dauerhafte Schlussfolgerungen im Prozess der Wissensgenerierung zu produzieren. Probanden mit einem niedrigen Skalenwert

¹⁸siehe auch die Schlussdiskussion dieser Arbeit, in der die theoretische Annahme, dass es sich beim DLC um ein emotional-motivationales Konstrukt handelt, diskutiert wird.

4. Korrelation mit Persönlichkeitsvariablen

tendieren dazu, eine Schlussfolgerung hinauszuzögern und eine einmal getroffene Schlussfolgerung leicht in Frage zu stellen. Die Skala besteht aus 16 Aussagen. Ein Frage ist z. B. „Ich mag es nicht, wenn die Aussage einer Person mehrdeutig ist“ oder „Wenn ich ein Problem lösen muss, verschwende ich im Allgemeinen keine Zeit damit, die unterschiedlichen Standpunkte dazu zu erwägen.“ Die Probanden können auf einer sechsstufigen Skala den Aussagen unterschiedlich stark zustimmen. Die in der Studie von Schlink und Walther (2007, S. 156) beobachtete interne Konsistenz ist mit einem Cronbach's α von .78 hoch.

Da Probanden mit einem hohen NFCC dazu tendieren, schnell zu einer Lösung kommen zu wollen, sollten sie in der Pferdewette weniger Cues aufdecken. Es wird somit eine negative Korrelation zwischen dem DLC und dem Skalenwert des 16-NCCS erwartet.

Die Ergebnisse zeigen über die Studien hinweg signifikant positive *und* negative Korrelationen zwischen dem DLC und dem 16-NCCS; siehe Tabelle 4.1 (S. 140). Im Mittel über neun Studien liegt die Korrelation bei Pearson's $r = 0$, die Hypothese kann nicht bestätigt werden.¹⁹

Der regulatorische Fokus (Werth & Förster, 2007) besteht aus dem *Promotion* Fokus und dem *Prevention* Fokus. Probanden mit hohem *Promotion* Fokus konzentrieren sich auf die „Maximierung positiver Ergebnisse“ und das „Erreichen von Idealen und Gewinnen“ (Werth & Förster, 2007, S. 33). Probanden mit hohem *Prevention* Fokus konzentrieren sich auf das „Vermeiden von Verlusten“ (Werth & Förster, 2007, S. 33). Mit dem regulatorischen Fokus ist ein bestimmtes Verhalten verknüpft. So tendieren Probanden mit einem hohen *Promotion* Fokus „riskantes“ Verhalten zu zeigen und „Schnelligkeit auf Kosten der Genauigkeit“ zu maximieren, während Probanden mit einem hohen *Prevention* Fokus sich „konservativer“, „vorsichtiger“ geben und „Genauigkeit auf Kosten der Schnelligkeit“ maximieren (Werth & Förster, 2007, S. 34, Tabelle 2).

Die beiden Subskalen umfassen jeweils 6 Fragen für den *Promotion* Fokus und 5 Fragen für den *Prevention* Fokus, zu denen die Probanden auf einer fünfstufigen Skala angeben, wie häufig bestimmte Ereignisse

¹⁹Die chronologisch gesehen erste Studie der Dissertation - das ist die Mindset Priming Studie - erbrachte eine hohe positive Korrelation (Pearson's $r = .50$), woraufhin die beiden situationalen NFCC Studien, die in der Arbeit berichtet wurden, folgten, um das Bild zu dem Persönlichkeitsbefund zu komplettieren. Unter Berücksichtigung der Ergebnisse über die Studie hinweg ist zu überlegen, inwieweit das NFCC (zumindest) im Kontext des DLC ein theoretisch gewinnbringendes Konstrukt ist.

in ihrem Leben eingetroffen sind. Eine Frage zum *Promotion* Fokus ist z. B. „Wie häufig haben Sie es erlebt, dass Ihnen etwas gelang und Sie danach noch enthusiastischer an der Sache gearbeitet haben“; eine Frage zum *Prevention* Fokus ist z. B. „Haben Sie als Heranwachsender jemals Dinge getan, die Ihre Eltern verboten hatte?“ (Frage muss invertiert werden). Die Skala ist nicht publiziert, so dass keine Normwerte vorliegen.²⁰

Da Probanden mit einem hohen *Promotion* Fokus dazu tendieren, riskoreiches Verhalten zu zeigen, sollten sie in der Pferdewette weniger Cues aufdecken; da Probanden mit einem hohen *Prevention* Fokus dazu tendieren, risikoarmes Verhalten zu zeigen, sollten sie in der Pferdewette mehr Cues aufdecken. Als Index wurde die Differenz zwischen dem *Promotion* Fokus und dem *Prevention* Fokus berechnet. Es wird somit eine negative Korrelation zwischen dem DLC und dem Differenzwert erwartet.

Die Ergebnisse zeigen über die Studien hinweg keinen signifikanten Zusammenhang zwischen dem DLC und dem Differenzwert; siehe Tabelle 4.1 (S. 140). Im Mittel über sechs Studien liegt die Korrelation bei Pearson's $r = 0$, die Hypothese kann nicht bestätigt werden.

Die *Achievement Motives Scale* (AMS-R) (Lang & Fries, 2006) misst die beiden Komponenten der Leistungsmotivation, das ist die Furcht vor Misserfolg und die Hoffnung auf Erfolg.

Probanden mit einer hohen Furcht vor Misserfolg tendieren dazu, die möglichen negativen Konsequenzen einer gescheiterten Zielerreichung zu fokussieren. Probanden mit einer hohen Hoffnung auf Erfolg tendieren dazu, die möglichen positiven Konsequenzen einer gelungenen Zielerreichung zu fokussieren. Die beiden Subskalen umfassen jeweils 5 Aussagen, zu denen die Probanden auf einer vierstufigen Skala ihren Grad der Zustimmung angeben sollen. Eine Aussage zur Furcht vor Misserfolg ist z. B. „Es beunruhigt mich, etwas zu tun, wenn ich nicht sicher bin, dass ich es kann.“; eine Aussage zur Hoffnung auf Erfolg ist z. B.: „Mich reizen Situationen, in denen ich meine Fähigkeiten testen kann“. Die von Lang und Fries (2006, Table 1, S. 220) berichteten interne Konsistenzen sind mit Cronbach's α Werten $> .7$ hoch.

Da Probanden mit einer hohen Furcht vor Misserfolg möglicherweise dazu tendieren, sich risikoaversiver zu verhalten und somit in der

²⁰Die Skala wurde mir aus dem Labor von Dr. Jens Förster am 10.01.2008 per Email zugesendet mit der Information, dass die Skala „in unseren Studien immer sehr verlässlich“ war. Die interne Konsistenz der Subskalen ist mit einem Cronbach's $\alpha > .6$ in allen hier berichteten Studien zufriedenstellend.

4. Korrelation mit Persönlichkeitsvariablen

Pferdwette mehr Cues aufdecken, wird ein positiver Zusammenhang zwischen der Furcht vor Misserfolg und dem DLC erwartet. Zur Hoffnung auf Erfolg gibt es keine spezifische Hypothese, der Zusammenhang der Skala mit dem DLC wird explorativ untersucht.

Die Ergebnisse zeigen über die Studien hinweg keinen signifikanten Zusammenhang zwischen dem DLC und der Furcht vor Misserfolg; siehe Tabelle 4.1 (S. 140). Im Mittel über drei Studien liegt die Korrelation bei Pearson's $r = -.03$, die Hypothese kann nicht bestätigt werden.

Das NEO-Fünf Faktoren Inventar (Borkenau & Ostendorf, 1993) misst Neurotizismus, Extraversion, Offenheit für Erfahrungen, Gewissenhaftigkeit und Verträglichkeit.

Probanden²¹ mit einem hohen Neurotizismuswert tendieren dazu, ängstlich und emotional instabil zu sein, mit einem hohen Extraversionenswert, gesellig und optimistisch zu sein, mit einem hohen Wert in Offenheit für Erfahrungen, wissbegierig und interessiert zu sein, mit einem hohen Gewissenhaftigkeitswert, ordentlich und zuverlässig zu sein, und mit einem hohen Verträglichkeitswert, verständnisvoll und altruistisch zu sein; siehe auch Amelang und Zielinski (2002, S. 307). Jede Subskala beinhaltet 12 Aussagen, zu denen die Probanden auf einer fünfstufigen Skala ihren Grad der Zustimmung angeben sollen. Eine Aussage zu Neurotizismus lautet z. B. „Ich fühle mich oft angespannt und nervös“, eine Aussage zu Extraversion lautet z. B. „Ich habe gern viele Leute um mich herum“, eine Aussage zu Offenheit für Erfahrung lautet z. B. „Ich probiere oft neue und fremde Speisen aus“, eine Aussage zu Gewissenhaftigkeit lautet z. B. „Ich arbeite hart, um meine Ziele zu erreichen“ und eine Aussage zu Verträglichkeit lautet z. B. „Ich versuche stets rücksichtsvoll und sensibel zu handeln“. Die von Borkenau und Ostendorf (1993) berichteten interne Konsistenzen sind mit Cronbach's α Werten $> .7$ hoch.

Da Probanden mit einem hohen Neurotizismuswert dazu tendieren, ängstlich und möglicherweise übervorsichtig zu reagieren, werden Probanden mit hohen Neurotizismuswerten tendenziell mehr Cues aufdecken und somit ein höheres DLC haben. Es wird also eine positive Korrelation zwischen dem DLC und den Neurotizismuwerten erwartet. Da Probanden mit einem hohen Wert in Gewissenhaftigkeit dazu ten-

²¹Die Daten wurden in einer Zusatzstudie, in der der Einfluss des DLC auf eine Gruppendiskussion untersucht wird, erhoben und ist zum jetzigen Zeitpunkt noch nicht abgeschlossen und wird daher in dieser Arbeit nicht berichtet. Die Erhebung des DLC und der NEO-FFI Fragebögen fand in einer separaten Sitzung vor der Gruppendiskussion statt.

dieren, diszipliniert und möglicherweise „penibel“ (Amelang & Zielinski, 2002, S. 307) zu arbeiten, werden Probanden mit hohen Werten in Gewissenhaftigkeit tendenziell mehr Cues aufdecken und somit ein höheres DLC haben. Es wird also eine positive Korrelation zwischen dem DLC und den Werten in Gewissenhaftigkeit erwartet. Zu den anderen drei Skalen bestehen keine konkreten Hypothesen, ihr Zusammenhang mit dem DLC wird explorativ untersucht.

Es zeigt sich kein signifikanter Zusammenhang zwischen dem DLC und dem Neurotizismuswert und der Gewissenhaftigkeit; siehe Tabelle 4.1 (S. 140). Die Korrelationen liegen bei Pearson's $r = .06$ bzw. $.07$, die beiden Hypothesen können nicht bestätigt werden. Es zeigt sich jedoch ein signifikant positiver Zusammenhang zwischen dem DLC und der Offenheit für Erfahrung. Es ist jedoch fraglich, wie stabil der Befund ist (siehe die instabilen Befunde zum 16-NCCS weiter oben).

Die Nullbefunde für die Korrelation zwischen dem DLC und den erhobenen Persönlichkeitsmerkmalen setzen die Tradition der Nullbefunde für die Korrelation zwischen Persönlichkeit und Entscheidungsverhalten fort (Bröder, 2003, 2005). Es ist möglich, dass das DLC kognitiv-strategische Prozesse widerspiegelt, die sich nicht bzw. nicht stark durch motivationale Zustände beeinflussen lassen. Das ist jedoch eine hoch spekulative Hypothese, die auch dem DLC zugrundeliegenden motivationalen Konstrukt des *Anspruchsniveaus an gewünschter Urteilssicherheit* (Hausmann-Thürig, 2004) entgegenläuft.

Tabelle 4.1. Korrelationen (Pearson's r) zwischen dem DLC und dem *Bedürfnis nach kognitiver Geschlossenheit (16-NCCS)* (Schlink & Walther, 2007), der Differenz des *Promotion* und *Prevention* Fokus (Werth & Förster, 2007), der *Achievement Motives Scale (AMS-R)* (Lang & Fries, 2006) und den *NEO Fünf Faktoren Inventar (NEO-FFI)* der Persönlichkeit (Borkenau & Ostendorf, 1993).

| Studien | n | 16-NCCS | Promotion - Prevention | Achievement Motives Scale | | NEO-FFI | | | | |
|-------------------------|--------------------------------|---------|------------------------------|------------------------------|--------------------|---------|------|------|------|-------|
| | | | | Hoffn. Erfolg | Furcht Misserf. | Neur. | Ex. | Off. | Gew. | Vert. |
| 1 Reliabilität | 51 | -.33* | -.11 | - | - | - | - | - | - | - |
| 2 Oberfläche | 51 | .06 | -.14 | .11 | -.11 | - | - | - | - | - |
| 3 Konzept. P. | 60 | -.34* | .24 | - | - | - | - | - | - | - |
| 4 Mindset P. | 39 | .50* | - | - | - | - | - | - | - | - |
| 5 NFCC Attraktivität | 40 | .24 | .05 | - | - | - | - | - | - | - |
| 6 NFCC Lärm | 71 | 0 | .09 | -.05 | 0 | - | - | - | - | - |
| 7 SDT Kat. | 48 | 0 | -.10 | 0 | 0 | - | - | - | - | - |
| 8 SDT Rekog. | 52 | - | - | - | - | - | - | - | - | - |
| Zusatzstudie | 107 (16-NCCS) 110 (NEO-FFI) | -.14 | - | - | - | .06 | -.12 | .27* | .07 | -.01 |
| Mittelwert | - | 0 | 0 | .02 | -.03 | - | - | - | - | - |
| Streuung | - | .28 | .14 | .08 | .06 | - | - | - | - | - |

Anmerkungen: * $p < .05$.

5. Diskussion

70.91% der Probanden folgen im Mittel über alle Studien hinweg dem DLC Modell des Abbruchs der Informationssuche in der Testphase der Pferdewette. Der Abbruch der Informationssuche lässt sich somit über das DLC für die Mehrheit der Probanden modellieren.

Das DLC wird in der Messaufgabe der Pferdewette reliabel erfasst und nicht durch irrelevante Merkmale der Messaufgabe - das sind die Merkmale, die irrelevant für das Messmodell sind - beeinflusst (Studie 1). Das DLC ist in einem befriedigenden Ausmaß zeitlich stabil (Pearson's $r = .61$), was für ein Persönlichkeitskonstrukt spricht, und hängt mit der Anzahl der konsultierten Cues in einer unabhängigen zweiten Entscheidungsaufgabe zusammen (Studie 2).

Die Höhe des DLC kann weder durch ein konzeptuelles Priming (Studie 3) noch durch ein Mindset Priming (Studie 4) beeinflusst werden. Die Höhe des DLC kann weder durch eine hohe Attraktivität der Folgeaufgabe (Studie 5) noch durch einen störenden Ton in der Erhebungssituation (Studie 6), beides Manipulationen des situationalen NFCC, verändert werden. Das DLC lässt sich also nicht - zumindest für die hier untersuchten Konstrukte - situational beeinflussen.

Das DLC hängt nicht mit dem Entscheidungskriterium der Signalentdeckungstheorie in einer Kategorisierungsaufgabe (Studie 7) und in einer Rekognitionsaufgabe (Studie 8) zusammen. Das DLC hängt nicht mit Persönlichkeitskonstrukten aus der Motivationspsychologie und den Big Five der Persönlichkeit zusammen, bis auf die berichtete geringe positive Korrelation (Pearson's $r = .27$) zwischen DLC und Offenheit für Erfahrung.

Für diese Reihe an Nullbefunden kann es statistische, konzeptuelle und methodische Gründe geben.

5.1. Statistische Gründe

Die Power der statistischen Testungen lag im Mittel über alle Studien bei .60 (SD = .21). Eine geringe Teststärke bedeutet einen erhöhte

Wahrscheinlichkeit für einen β -Fehler. Es ist möglich, dass vorhandene Effekte keine „Chance“ hatten, detektiert zu werden. Es ist somit für (mögliche) Replikationsstudien notwendig, größere Stichproben zu erheben.

5.2. Konzeptuelle Gründe

Die Evidenz der Studien dieser Arbeit spricht gegen die psychologische Bedeutsamkeit des DLC. Das DLC soll ein Anspruchsniveau an gewünschter Urteilssicherheit (Hausmann-Thürig, 2004) abbilden. Die Studien geben jedoch keinen Hinweis dafür, dass das DLC ein emotional-motivationales Konstrukt ist. Das DLC lässt sich weder durch die situationale Manipulationen theoretisch verwandter Konstrukte beeinflussen, noch lassen sich bedeutsame Zusammenhänge mit verwandten Persönlichkeitsvariablen aufzeigen. Es ist also möglich, dass das DLC Modell nur ein zeitstabiles Messwerkzeug ist, um den Abbruch der Informationssuche in probabilistischen Entscheidungsaufgaben, wie sie in den Studien verwendet wurden, zu erfassen. Das DLC wäre dann über den Kontext der in den Studien verwendeten probabilistischen Entscheidungsaufgaben (Pferdewette und Börsenspiel) hinaus unbedeutsam.

Es ist weiterhin möglich, dass das DLC ein stabiles psychologisch bedeutsames Persönlichkeitsmerkmal ist, das nicht von situationalen Bedingungen, die im Labor hergestellt werden können, beeinflusst wird. Das kann daran liegen, dass es sich bei dem DLC um einen chronischen Motivationszustand handelt, der nicht (bzw. wenig) mit der Situation interagiert: Das wäre aber in einer Entscheidungsumwelt, in der nicht jede Entscheidung die gleiche Bedeutsamkeit besitzt und nach einer unterschiedlich gründlichen Informationssuche verlangt, wenig adaptiv. Es ist weiterhin möglich, dass die situationalen Manipulationen im Labor zu schwach waren, um das DLC zu beeinflussen. So konnten Bröder und Schiffer (2006a) zeigen, dass Probanden dazu neigen, eine einmal verwendete Strategie beizubehalten, selbst wenn die Umwelt sich verändert und die Strategie nicht länger adaptiv ist; siehe zum Thema der Routine-Effekte S. 44 in Kapitel 1.2.5. Ein weiterer möglicher Grund ist, dass das DLC kein emotional-motivationales Konstrukt abbildet, sondern eher mit kognitiv-strategischen Konstrukten verwandt ist, die in den Studien dieser Arbeit nicht untersucht wurden. So konnte Hausmann-Thürig (2004) zeigen, dass Probanden ihr Stoppverhalten den Suchkosten der Cues anpassen: Kosten-Nutzen Abwägungen haben so-

mit durchaus einen Einfluss auf das DLC. In zukünftige Studien sollten somit emotional-motivationale und kognitiv-strategische Komponenten erfasst werden, um die Prädiktionskraft beider Variablen für die Varianz der Probanden im DLC zu bestimmen.

5.3. Methodische Gründe

In der verwendeten Methode zur Messung des DLC interessiert nur, ob ein Proband nach *einem* Cue in Abhängigkeit zur Validität des Cues die Suche nach weiteren Cues abbricht. Wie viele Cues er *nach* dem zuerst präsentierten Cue aufdeckt, wie die Validitäten dieser Cues den Abbruch nach dem ersten Cue beeinflussen und wie die Cues zu einer Entscheidung integriert werden, wird im DLC nicht abgebildet. Dass aus dem DLC keine Hypothese zur Integration von Cues zu einer Entscheidung abgeleitet werden kann, ist erst einmal nicht problematisch, wenn in dem Modell "nur" der Abbruch der Informationssuche abgebildet werden soll. Es ist jedoch hochproblematisch, dass in der Berechnung des DLC die Validitäten weiterer Cues, die ein Proband aufdeckt, wenn er sich mit der Validität des ersten Cues nicht zufrieden gibt, ignoriert werden. So ist die Varianz zwischen Probanden im Abbruchverhalten in Relation zu den Validitäten der Cues, wenn *mehr* als ein Cue konsultiert wird, sicherlich informativ für ein Anspruchsniveau an gewünschter Urteilssicherheit. Es erscheint somit sinnvoll, das DLC Modell dahingehend zu erweitern, ein DLC basierend auf allen Validitäten der von den Probanden konsultierten Cues zu bestimmen.

Eine Erweiterung des DLC Modells um die aufgedeckten Cues nach dem ersten Cue führt zu einer theoretischen Frage: Es ist dann notwendig zu klären, wie Probanden Validitäten von Cues zur Bestimmung der Evidenz für eine Entscheidungsoption verrechnen. Diese (anspruchsvolle) Frage nach dem/den zugrundeliegenden Prozess(en) der Verrechnung kann jedoch auch erst einmal zurückgestellt werden und stattdessen ein (psychologisch weniger plausibler) Algorithmus getestet werden wie ihn die Autoren des im Theorieteil der Arbeit (S. 58) vorgestellten Evidenzakkumulations-Modell vorschlagen (Lee & Cummins, 2004): Im Modell wird die Integration von Cues zu einem Urteil über eine (naive) Bayesianische Verrechnung der Validitäten modelliert.

Letztendlich ähneln sich das Evidenzakkumulations-Modell von Lee und Cummins (2004) und das DLC Modell in der Berechnung der Schwelle, die für einen Abbruch der Informationssuche erreicht wer-

5. Diskussion

den muss. Es gibt jedoch einen entscheidenden Unterschied: Das DLC wird über das Verhalten des Abbruchs der Informationssuche der Probanden berechnet. Ein optimales DLC liegt so, dass die Vorhersage des Abbruchs mit dem beobachteten Abbruch einhergeht: Liegt die Validität des (ersten) Cues über dem DLC, sollte bei modellerwartungstreuem Entscheidungsverhalten die Suche abgebrochen werden, liegt sie unterhalb des DLC, sollte die Suche fortgeführt werden. Im Evidenzakkumulations-Modell dagegen wird die Schwelle über die getroffenen Entscheidungen berechnet. Eine optimale Schwelle liegt so, dass die Wahlen über ein Set von Aufgaben, in denen die Cue-Muster und Cue-Validitäten unterschiedlich sind, möglichst gut vorhergesagt werden können. Im DLC wird somit das Prozessmaß des Abbruchs der Informationssuche als Indiz für eine Schwelle verwendet, während im Evidenzakkumulations-Modell das Outcome-Maß der Wahlen zur Berechnung einer Schwelle herangezogen wird.

Trotz dieses Unterschieds zwischen den Modellen in der Berechnung einer Evidenz-Schwelle, lässt sich das Evidenzakkumulations-Modell von Lee und Cummins (2004) prinzipiell auf die Probandendaten der Studien dieser Arbeit anwenden: Über alle Studien hinweg wurden die präsentierten Cue-Muster und Cue-Validitäten zusammen mit dem Stopp- und Entscheidungsverhalten der Probanden aufgezeichnet. Um das Evidenzakkumulations-Modell auf die Daten anzuwenden, wäre es jedoch notwendig, das Modell von dem zwei Optionen Fall in den Studien von Lee und Cummins (2004) auf den vier Optionen Fall der Pferdewette zu generalisieren. Es muss weiterhin überprüft werden, ob die Muster und Cue-Validitäten für einzelne Probanden überhaupt diagnostisch für bestimmte Schwellenausprägungen im Evidenzakkumulations-Modell sind: Eine diagnostische Aussage (d. h. über interindividuelle Unterschiede) über eine Schwelle kann nur gemacht werden, wenn die Verwendung eines (bzw. weniger) Cues zu einer unterschiedlichen Modellvorhersage über die Wahlen führt als die Integration mehrerer (bzw. aller) Cues. Da die Validitäten des zweiten bis siebten Cues in der Pferdewette jedoch per Zufall eingespielt wurden und somit nicht a-priori für eine optimale Modelltestung konstruiert wurden, müssen sie nicht per se diagnostisch für verschiedene Schwellen sein (Jekel et al., 2011). Abgesehen von diesen möglichen Schwierigkeiten, ist es jedoch prinzipiell möglich, neben dem hier validierten DLC Modell das Evidenzakkumulations-Modell von Lee und Cummins (2004) unter Berücksichtigung der Validitäten der konsultierten Cues und der getroffenen Wahlen in der Pferdewette anzuwenden und auf diesem Wege die Daten

zu reanalysieren.

5.4. Fazit

Die Evidenz zur Validierung des DLC kann in einem ersten Fazit folgendermaßen zusammengefasst werden: Die Messaufgabe der Pferdewette erfasst das DLC reliabel. Das DLC ist weiterhin ein zeitstabiles Konstrukt und erfüllt somit eine Bedingung für ein Persönlichkeitskonstrukt. Die Frage nach der Validität des DLC, d. h. die Frage nach der konzeptuellen Bedeutung des DLC, konnte nicht geklärt werden. Zwei zentrale Gründe dafür wurden in der Diskussion genannt: Es ist möglich, dass die Schwelle nicht ein Anspruchsniveau an gewünschter Urteilssicherheit abbildet. Es ist möglich, dass das Messmodell in seiner jetzigen Form unzureichend ist. Das DLC vernachlässigt wichtige Elemente der Entscheidungssituation, das ist das Abbruchverhalten der Informationssuche bei mehr als einem Cue. Eine Erweiterung des DLC Modells und ein mögliches Alternativmodell wurden diskutiert. Es ist eine empirische Frage, ob die Einbeziehung mehrerer Cues in die Bestimmung einer Abbruchschwelle zu erhöhter bedeutsamer Varianz im DLC führt, die dann möglicherweise sensitiv(er) für die in der Arbeit verwendeten situationalen Bedingungen ist und bedeutsame Zusammenhänge zu den erhobenen Persönlichkeitsvariablen aufweist.

Literatur

- Aarts, H. & Dijksterhuis, A. (2003). The silence of the library: Environment, situational norm, and social behavior. *Journal of Personality and Social Psychology*, *84*, 18–28.
- Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, *19*, 716–723.
- Amelang, M. & Zielinski, W. (2002). *Psychologische Diagnostik und Intervention* (3. Aufl.). Heidelberg: Springer.
- Baddeley, A., Eysenck, M. W. & Anderson, M. (2009). *Memory*. East Sussex: Psychology Press.
- Bargh, J. A. & Chartrand, T. L. (2000). The mind in the middle: A practical guide to priming and automaticity research. In H. T. Reis & C. M. Judd (Hrsg.), *Handbook of research methods in social and personality psychology* (S. 253–285). New York: Cambridge.
- Bargh, J. A., Chen, M. & Burrows, L. (1996). Automaticity of social behavior: Direct effects of trait construct and stereotype activation on action. *Journal of Personality and Social Psychology*, *71*, 230–244.
- Bargh, J. A. & Williams, E. L. (2006). The automaticity of social life. *Current Directions in Psychological Science*, *15*, 1–4.
- Baschek, I.-L., Bredenkamp, J., Oehrle, B. & Wippich, W. (1977). Bestimmung der Bildhaftigkeit (I), Konkretheit (C) und der Bedeutungshaltigkeit (m') von 800 Substantiven. *Zeitschrift für experimentelle und angewandte Psychologie*, *24*, 353–396.
- Baschek, I.-L., Bredenkamp, J., Oehrle, B. & Wippich, W. (1994). Bestimmung der Bildhaftigkeit, Konkretheit und der Bedeutungshaltigkeit von 800 Substantiven. In W. Hager & M. Hasselhorn (Hrsg.), *Handbuch deutschsprachiger wortnormen* (S. 174–186). Göttingen, Hogrefe.
- Bayes, T. (1763). An essay towards solving a problem in the doctrine of chances. *Philosophical Transactions of the Royal Society*, *53*, 376–398.
- Bernoulli, D. (1738). Specimen theoriae novae de mensura sortis (Exposition of a new theory on the measurement of risk, Übersetzung 1954). *Econometrica*, *22*, 23–36.
- Borges, B., Goldstein, D. G., Ortman, A. & Gigerenzer, G. (1999). Can ignorance beat the stock market. In G. Gigerenzer, P. M. Todd & The ABC Research Group (Hrsg.), *Simple heuristics that make us smart* (S. 59–72). Cambridge: University Press.
- Borkenau, P. & Ostendorf, F. (1993). *NEO-Fünf-Faktoren-Inventar (NEO-FFI)*. Göttingen: Hogrefe.
- Boyd, M. (2001). On ignorance, intuition, and investing: A bear market test of the recognition heuristic. *The Journal of Psychology and Financial Markets*, *2*, 150–156.
- Brandstätter, E., Gigerenzer, G. & Hertwig, R. (2006). The priority heuristic: Making choices without trade-offs. *Psychological Review*, *113*, 409–432.
- Bröder, A. (2000a). A methodological comment on behavioral decision

LITERATUR

- research. *Psychologische Beiträge*, 42, 645–662.
- Bröder, A. (2000b). Assessing the empirical validity of the “Take-The-Best” heuristic as a model of human probabilistic inference. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 26, 1332–1346.
- Bröder, A. (2000c). “Take The Best - Ignore The Rest”: Wann entscheiden Menschen begrenzt rational? Berlin: Pabst.
- Bröder, A. (2003). Decision making with the ‘adaptive toolbox’: Influence of environmental structure, intelligence, and working memory load. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 29, 611–625.
- Bröder, A. (2005). *Entscheiden mit der “adaptiven Werkzeugkiste”*: Ein empirisches Forschungsprogramm. Lengerich: Pabst.
- Bröder, A. (2007). The Quest for Take The Best - Insights and Outlooks from Experimental Research.
- Bröder, A. (2009). Cues. In M. Kattan (Hrsg.), *Encyclopedia of medical decision making* (Bd. 1, S. 245–247). London: Sage.
- Bröder, A. (2010). Outcome-based strategy classification. In A. Glöckner & C. L. M. Wittemann (Hrsg.), *Foundations for tracing intuition: Challenges and methods* (S. 61–82). London: Psychology Press & Routledge.
- Bröder, A. (2011). The quest for Take The Best: Insights and outlooks from experimental research. Manuskript im Druck.
- Bröder, A. & Newell, B. R. (2008). Challenging some common beliefs: Empirical work within the adaptive toolbox metaphor. *Judgment and Decision Making*, 3, 205–214.
- Bröder, A., Newell, B. R. & Platzer, C. (2010). Cue integration vs. exemplar-based reasoning in multi-attribute decisions from memory: A matter of cue representation. *Judgment and Decision Making*, 5, 326–338.
- Bröder, A. & Schiffer, S. (2003a). Bayesian strategy assessment in multi-attribute decision making. *Journal of Behavioral Decision Making*, 16, 193–213.
- Bröder, A. & Schiffer, S. (2003b). Take The Best versus simultaneous feature matching: Probabilistic inferences from memory and effects of representation format. *Journal of Experimental Psychology: General*, 132, 277–293.
- Bröder, A. & Schiffer, S. (2006a). Adaptive flexibility and maladaptive routines in selecting fast and frugal decision strategies. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 32, 904–918.
- Bröder, A. & Schiffer, S. (2006b). Stimulus format and working memory in fast and frugal strategy selection. *Journal of Behavioral Decision Making*, 19, 361–380.
- Bröder, A. & Schütz, J. (2009). Recognition ROCs are curvilinear or are they? On premature arguments against the two-high-threshold model of recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 35, 587–606.
- Browne, M. W. (2000). Cross-validation methods. *Journal of Mathematical Psychology*, 44, 108–132.
- Brunswik, E. (1955). Representative design and probabilistic theory in a functional psychology. *Psychological Review*, 62, 193–217.
- Brunswik, E. (1956). *Perception and the representative design of psychological experiments* (2. Aufl.).

- Berkeley: University of California Press.
- Burnham, K. P. & Anderson, D. R. (2002). *Model selection and multimodel inference: A practical information-theoretic approach* (2. Aufl.). New York: Springer.
- Busemeyer, J. & Townsend, J. (1993). Decision field theory: A dynamic-cognitive approach to decision making in an uncertain environment. *Psychological Review*, *100*, 432–459.
- Chaiken, S., Liberman, A. & Eagly, A. H. (1989). Heuristic and systematic information processing within and beyond the persuasion process. In J. S. Uleman & J. A. Bargh (Hrsg.), *Unintended thought* (S. 212–252). New York: Guilford Press.
- Chaiken, S. & Trope, Y. (1999). *Dual-process theories in social psychology*. New York: Guilford Press.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2. Aufl.). Hillsdale, NJ: Erlbaum.
- Cosmides, L. & Tooby, J. (1994). Beyond intuition and instinct blindness: Toward an evolutionarily rigorous cognitive science. *Cognition*, *50*, 41–77.
- Cutting, J. E. (2000). Accuracy, scope, and flexibility of models. *Journal of Mathematical Psychology*, *44*, 3–19.
- Czerlinski, J., Gigerenzer, G. & Goldstein, D. G. (1999). How good are simple heuristics. In G. Gigerenzer, P. M. Todd & The ABC Research Group (Hrsg.), *Simple heuristics that make us smart* (S. 97–118). New York: Oxford University Press.
- Dawes, R. M. (1979). The robust beauty of improper linear models in decision making. *American Psychologist*, *34*, 571–582.
- DePaulo, B. M., Lindsay, J. J., Malone, B. E., Muhlenbruck, L., Charlton, K. & Cooper, H. (2003). Cues to deception. *Psychological Bulletin*, *129*, 74–118.
- Devine, P. G. (1989). Stereotypes and prejudice: Their automatic and controlled components. *Journal of Personality and Social Psychology*, *56*, 5–18.
- Dhimi, M. K., Hertwig, R. & Hoffrage, U. (2004). The role of representative design in an ecological approach to cognition. *Psychological Bulletin*, *130*, 959–988.
- Dijksterhuis, A. & Bargh, J. A. (2001). The perception-behavior expressway: Automatic effects of social perception on social behavior. In M. P. Zanna (Hrsg.), *Advances in Experimental Social Psychology* (Bd. 33, S. 1–40). San Diego: Academic Press.
- Dougherty, M. R. P., Franco-Watkins, A. M. & Thomas, R. (2008). Psychological plausibility of the theory of probabilistic mental models and the fast and frugal heuristics. *Psychological Review*, *115*, 199–213.
- Dougherty, M. R. P., Gettys, C. F. & Ogden, E. E. (1999). MINERVA-DM: A memory processes model for judgments of likelihood. *Psychological Review*, *106*, 180–209.
- Eid, M., Gollwitzer, M. & Schmitt, M. (2010). *Statistik und Forschungsmethoden*. Weinheim: Beltz.
- Epley, N. & Gilovich, T. (1999). Just going Along: Nonconscious priming and conformity to social pressure. *Journal of Experimental Social Psychology*, *35*, 578–589.

LITERATUR

- Fiedler, K. & Walka, I. (1993). Training lie detectors to use nonverbal cues instead of global heuristics. *Human Communication Research, 20*, 199–223.
- Fishbach, A., Dhar, R. & Zhang, Y. (2006). Subgoals as substitutes or complements: The role of goal accessibility. *Journal of Personality and Social Psychology, 91*, 232–242.
- Förster, J. & Liberman, N. (2007). Knowledge activation. In A. W. Kruglanski & E. T. Higgins (Hrsg.), *Social psychology: Handbook of basic principles* (2. Aufl., S. 201–231). New York: The Guilford Press.
- Friedman, R. S., Fishbach, A., Förster, J. & Werth, L. (2003). Attentional priming effects on creativity. *Creativity Research Journal, 15*, 277–286.
- Garcia-Retamero, R. & Rieskamp, J. (2008). Adaptive mechanisms for treating missing information: A simulation study. *58*, 547–568.
- Garcia-Retamero, R. & Rieskamp, J. (2009). Do people treat missing information adaptively when making inferences? *Quarterly Journal of Experimental Psychology, 62*, 1991–2013.
- Gigerenzer, G. (1996). On narrow norms and vague heuristics: A reply to Kahneman and Tversky (1996). *Psychological Review, 103*, 592–596.
- Gigerenzer, G. & Brighton, H. (2009). Homo heuristicus: Why biased minds make better inferences. *Topics in Cognitive Science, 1*, 107–143.
- Gigerenzer, G., Czerlinski, J. & Martignon, L. (1999). How good are fast and frugal heuristics? In J. Shanteau, B. A. Mellers & D. A. Schum (Hrsg.), *Decision science and technology: Reflections on the contributions of Ward Edwards* (S. 81–103). Norwell: Kluwer Academic Publishers.
- Gigerenzer, G. & Gaissmaier, W. (2011). Heuristic decision making. *Annual Review of Psychology, 62*, 451–482.
- Gigerenzer, G. & Goldstein, D. G. (1996). Reasoning the fast and frugal way: Models of bounded rationality. *Psychological Review, 103*, 650–669.
- Gigerenzer, G. & Goldstein, D. G. (1999). Betting on one good reason. In G. Gigerenzer, P. M. Todd & The ABC Research Group (Hrsg.), *Simple heuristics that make us smart* (S. 75–95). New York: Oxford University Press.
- Gigerenzer, G. & Hoffrage, U. (1995). How to improve bayesian reasoning without instruction: Frequency formats. *Psychological Review, 102*, 684–704.
- Gigerenzer, G., Hoffrage, U. & Kleinbölting, H. (1991). Probabilistic mental models: A Brunswikian theory of confidence. *Psychological Review, 98*, 506–528.
- Gigerenzer, G. & Kurz, E. M. (2001). Vicarious functioning reconsidered: A fast and frugal lens model. In K. R. Hammond & T. R. Stewart (Hrsg.), *The essential Brunswik: Beginnings, explications, applications* (S. 342–357). Oxford: University Press.
- Gigerenzer, G. & Todd, P. M. (1999). Fast and frugal heuristics: The adaptive toolbox. In G. Gigerenzer, P. M. Todd & The ABC Research Group (Hrsg.), *Simple heuristics that make us smart* (S. 3–34). Oxford: University Press.
- Gigerenzer, G., Todd, P. M. & The ABC Research Group. (1999). *Simple heuristics that make us*

- smart*. New York: Oxford University Press.
- Glöckner, A. (2006). *Automatische Prozesse bei Entscheidungen: Das dominierende Prinzip menschlicher Entscheidungen. Intuition, komplex-rationale Analyse oder Reduktion?* Hamburg: Verlag Dr. Kovač.
- Glöckner, A. (2009). How Distinct are Intuition How Distinct are Intuition and Deliberation ? An Eye-Tracking Analysis of Instruction-Induced Decision Modes.
- Glöckner, A. (2010). Multiple measure strategy classification: Outcomes, decision times and confidence ratings. In A. Glöckner & C. L. M. Witteman (Hrsg.), *Foundations for tracing intuition* (S. 83–105). London: Psychology Press & Routledge.
- Glöckner, A. & Betsch, T. (2008a). Modeling option and strategy choices with connectionist networks: Towards an integrative model of automatic and deliberate decision making. *Judgment and Decision Making*, 3, 215–228.
- Glöckner, A. & Betsch, T. (2008b). Multiple-reason decision making based on automatic processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 34, 1055–1075.
- Glöckner, A. & Betsch, T. (2011). The empirical content of theories in judgment and decision making: Shortcomings and remedies. *Judgment and Decision Making*, 6, 711–721.
- Glöckner, A., Betsch, T. & Schindler, N. (2009). Coherence shifts in probabilistic inference tasks. *Journal of Behavioral Decision Making*.
- Glöckner, A. & Herbold, A.-K. (2011). An eye-tracking study on information processing in risky decisions: Evidence for compensatory strategies based on automatic processes. *Journal of Behavioral Decision Making*, 24, 71–98.
- Glöckner, A. & Witteman, C. L. M. (2010). *Foundations for tracing intuition*. London: Psychology Press & Routledge.
- Goldstein, D. G. & Gigerenzer, G. (1999). The recognition heuristic: How ignorance makes us smart. In G. Gigerenzer, P. M. Todd & The ABC Research Group (Hrsg.), *Simple heuristics that make us smart* (S. 37–58). Cambridge: University Press.
- Goldstein, D. G. & Gigerenzer, G. (2002). Models of ecological rationality: The recognition heuristic. *Psychological Review*, 109, 75–90.
- Gollwitzer, P. M., Heckhausen, H. & Steller, B. (1990). Deliberative and implemental mind-sets: Cognitive tuning toward congruous thoughts and information. *Journal of Personality and Social*, 59, 1119–1127.
- Gosling, S., Ko, S., Mannarelli, T. & Morris, M. (2002). A room with a cue: Personality judgments based on offices and bedrooms. *Journal of Personality and Social Psychology*, 82, 379–398.
- Green, D. M. & Swets, J. A. (1966). *Signal detection theory and psychophysics*. New York: Wiley.
- Grünwald, P. (2000). Model selection based on minimum description length. *Journal of Mathematical Psychology*, 44, 133–152.
- Grünwald, P. (2005). A tutorial introduction to the minimum description principle. In P. Grünwald, I. J. Myung & M. Pitt (Hrsg.),

LITERATUR

- Advances in minimum description length: Theory and applications* (S. 23–81). Cambridge: MIT Press.
- Hammond, K. R. (1966). *The psychology of Egon Brunswik*. New York: Holt, Rinehart Winston.
- Hammond, K. R. & Stewart, T. R. (2001). *The essential Brunswik: Beginnings, explications, applications*. Oxford: University Press.
- Hausmann, D. & Läge, D. (2008). Sequential evidence accumulation in decision making: The individual desired level of confidence can explain the extent of information acquisition Daniel. *Judgment and Decision Making*, 3, 229–243.
- Hausmann-Thürig, D. (2004). *Informationssuche im Entscheidungsprozess: Die Nützlichkeit von Hinweis-Cues und der Anspruch an Urteilssicherheit*. Zürich: Zentralstelle der Studentendruckerei.
- Hertwig, R., Davis, J. N. & Sulloway, F. J. (2002). Parental investment: How an equity motive can produce inequality. *Psychological Bulletin*, 128, 728–745.
- Hertwig, R., Herzog, S. M., Schooler, L. J. & Reimer, T. (2008). Fluency heuristic: A model of how the mind exploits a by-product of information retrieval. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 34, 1191–1206.
- Hertwig, R., Hoffrage, U. & Martignon, L. (1999). Quick estimation: Letting the environment do the work. In G. Gigerenzer, P. M. Todd & The ABC Research Group (Hrsg.), *Simple heuristics that make us smart* (S. 209–234). Cambridge: University Press.
- Hertwig, R. & Pleskac, T. (2008). The game of life: How small samples render choice simpler. In N. Chater & M. Oaksford (Hrsg.), *The probabilistic mind: Prospects for Bayesian cognitive science* (S. 209–235). Oxford: University Press.
- Higgins, E. T. & Chaires, W. M. (1980). Accessibility of interrelational constructs: Implications for stimulus encoding and creativity. *Journal of Experimental Social Psychology*, 16, 348–361.
- Higgins, E. T., Rholes, W. S. & Jones, C. R. (1977). Category accessibility and impression formation. *Journal of Experimental Social Psychology*, 13, 141–154.
- Hilbig, B. (2008). Individual differences in fast-and-frugal decision making: Neuroticism and the recognition heuristic. *Journal of Research in Personality*, 42, 1641–1645.
- Hilbig, B. E. (2010a). Precise models deserve precise measures: A methodological dissection. *Judgment and Decision Making*, 5, 272–284.
- Hilbig, B. E. (2010b). Reconsidering “evidence” for fast and frugal heuristics. *Psychonomic Bulletin & Review*, 17, 923–930.
- Hilbig, B. E., Erdfelder, E. & Pohl, R. F. (2010). One-reason decision making unveiled: A measurement model of the recognition heuristic. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 36, 123–134.
- Hilbig, B. E. & Richter, T. (2011). Homo heuristicus outnumbered: Comment on Gigerenzer and Brighton (2009). *Topics in Cognitive Science*, 3, 187–196.
- Hintzman, D. L. (1988). Judgments of frequency and recognition memory in a multiple-trace memory model. *Psychological Review*, 95, 528–551.

- Hoffrage, U. & Hertwig, R. (2006). Which world should be represented in representative design. In K. Fiedler & P. Juslin (Hrsg.), *Information sampling and adaptive cognition* (S. 381–408). Cambridge: University Press.
- Hogarth, R. M. & Karelaia, N. (2006). “Take-the-Best” and other simple strategies: Why and when they work “well” with binary cues. *Theory and Decision*, 61, 205–249.
- Hogarth, R. M. & Karelaia, N. (2007). Heuristic and linear models of judgment: Matching rules and environments. *Psychological Review*, 114, 733–758.
- Horstmann, N., Ahlgrimm, A. & Glöckner, A. (2009). How distinct are intuition and deliberation? An eye-tracking analysis of instruction-induced decision modes. *Judgment and Decision Making*, 4, 335–354.
- Huang, Y. & Wang, L. (2010). Sex differences in framing effects across task domain. *Personality and Individual Differences*, 48, 649–653.
- Jäger, A. O., Süß, H.-M. & Beauducel, A. (1997). *Berliner Intelligenzstruktur-Test*. Göttingen: Hogrefe.
- Jekel, M. (in Vorbereitung). Taking the role of Brunswik’s representative design in an ecological approach to cognition seriously: Inflated accuracy rates for decision strategies due to nonrepresentative sampling.
- Jekel, M., Fiedler, S. & Glöckner, A. (2011). Diagnostic task selection for strategy classification in judgment and decision making: Theory, validation, and implementation in R. *Judgment and Decision Making*, 6, 782–799.
- Jekel, M. & Glöckner, A. (in Vorbereitung). Doing justice to Benjamin Franklin’s ideas: Overestimation of heuristics caused by wrong implementations of weighted compensatory strategies.
- Jekel, M., Glöckner, A., Fiedler, S. & Bröder, A. (unter Begutachtung). The rationality of different kinds of intuitive processes.
- Jekel, M., Nicklisch, A. & Glöckner, A. (2010). Implementation of the multiple-measure maximum likelihood strategy classification method in R: Addendum to Glöckner (2009) and practical guide for application. *Judgment and Decision Making*, 5, 54–63.
- Jost, J. T., Glaser, J., Kruglanski, A. W. & Sulloway, F. J. (2003). Political conservatism as motivated social cognition. *Psychological Bulletin*, 129, 339–375.
- Jungermann, H., Pfister, H.-R. & Fischer, K. (2010). *Die Psychologie der Entscheidung* (3. Aufl.). Heidelberg: Springer.
- Juslin, P., Karlsson, L. & Olsson, H. (2008). Information integration in multiple cue judgment: A division of labor hypothesis. *Cognition*, 106, 259–298.
- Juslin, P. & Persson, M. (2002). PROBABILITIES from EXemplars (PROBEX): A “lazy” algorithm for probabilistic inference from generic knowledge. *Cognitive Science*, 26, 563–607.
- Kahneman, D., Slovic, P. & Tversky, A. (1982). *Judgment under uncertainty: Heuristics and biases*. Cambridge: University Press.
- Kahneman, D. & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica*, 47, 263–292.

LITERATUR

- Kahneman, D. & Tversky, A. (1996). On the reality of cognitive illusions. *Psychological Review*, *103*, 582–591.
- Kahneman, D. & Tversky, A. (2000). *Choices, values, and frames*. Cambridge: University Press.
- Karlsson, L., Juslin, P. & Olsson, H. (2008). Exemplar-based inference in multi-attribute decision making: Contingent, not automatic, strategy shifts? *Judgment and Decision Making*, *3*, 244–260.
- Katsikopoulos, K., Schooler, L. & Hertwig, R. (2010). The robust beauty of ordinary information. *Psychological Review*, *117*, 1259–1266.
- Krohne, H. W., Egloff, B., Kohlmann, C. W. & Tausch, A. (1996). Untersuchungen mit einer deutschen Version der „Positive and Negative Affect Schedule“ (PANAS). *Diagnostica*, *42*, 139–156.
- Kruglanski, A. W. (2004). *The psychology of closed mindedness*. New York: Psychology Press.
- Kruglanski, A. W. & Freund, T. (1983). The freezing and unfreezing of lay-inferences: Effects on impression primacy, ethnic stereotyping, and numerical anchoring. *Journal of Experimental Social Psychology*, *19*, 448–468.
- Kruglanski, A. W. & Thompson, E. P. (1999). Persuasion by a single route: A view from the unimodel. *Psychological Inquiry*, *10*, 83–109.
- Kruglanski, A. W. & Webster, D. M. (1991). Group members' reactions to opinion deviates and conformists at varying degrees of proximity to decision deadline and of environmental noise. *Journal of Personality and Social Psychology*, *61*, 212–225.
- Kruglanski, A. W. & Webster, D. M. (1996). Motivated closing of the mind: “Seizing” and “freezing”. *Psychological Review*, *103*, 263–283.
- Kruglanski, A. W., Webster, D. M. & Klem, A. (1993). Motivated resistance and openness to persuasion in the presence or absence of prior information. *Journal of Personality and Social Psychology*, *65*, 861–876.
- Kruschke, J. K. (2010). What to believe: Bayesian methods for data analysis. *Trends in cognitive sciences*, *14*, 293–300.
- Lang, J. W. B. & Fries, S. (2006). A revised 10-item version of the Achievement Motives Scale: Psychometric properties in german-speaking samples. *European Journal of Psychological Assessment*, *22*, 216–224.
- Lee, M. D. & Cummins, T. D. R. (2004). Evidence accumulation in decision making: Unifying the “take the best” and the “rational” models. *Psychonomic Bulletin & Review*, *11*, 343–352.
- Lewandowsky, S. (1993). The rewards and hazards of computer simulations. *Psychological Science*, *4*, 236–243.
- Lewandowsky, S. & Farrell, S. (2011). *Computational modeling in cognition: Principles and practice*. London: Sage.
- Lombardi, W. J., Higgins, E. T. & Bargh, J. A. (1987). The role of consciousness in priming effects on categorization: Assimilation versus contrast as a function of awareness of the priming task. *Personality and Social Psychology Bulletin*, *13*, 411–429.
- Lopes, L. (1991). The rhetoric of irrationality. *Theory & Psychology*, *1*, 65–82.
- Lorenz, K. (1967). *Über tierisches und menschliches Verhalten: Aus*

- dem Werdegang der Verhaltenslehre (Gesammelte Abhandlungen). Berlin: Deutsche Buchgemeinschaft.
- Lukas, J. (2006). Signalentdeckungstheorie. In J. Funke & P. A. Frensch (Hrsg.), *Handbuch der Psychologie; Bd. 5: Handbuch der Allgemeinen Psychologie - Kognition* (S. 732–740). Göttingen: Hogrefe.
- Macmillan, N. A. & Creelman, C. D. (2005). *Detection theory: A user's guide* (2. Aufl.). New Jersey: Lawrence Erlbaum.
- Macrae, C. N. & Johnston, L. (1998). Help, I need somebody: Automatic action and inaction. *Social Cognition, 16*, 400–417.
- Marewski, J. N., Gaissmaier, W., Schooler, L. J., Goldstein, D. G. & Gigerenzer, G. (2010). From recognition to decisions: Extending and testing recognition-based models for multialternative inference. *Psychonomic Bulletin & Review, 17*, 287–309.
- Martignon, L. (1999). Comparing fast and frugal heuristics and optimal models. In G. Gigerenzer & R. Selten (Hrsg.), *Bounded rationality: The adaptive toolbox* (S. 147–171). Cambridge: MIT Press.
- Martignon, L. & Hoffrage, U. (1999). Why does one-reason decision making work? In G. Gigerenzer, P. M. Todd & The ABC Research Group (Hrsg.), *Simple heuristics that make us smart* (S. 110–140). New York: Oxford University Press.
- Martignon, L. & Hoffrage, U. (2002). Fast, frugal, and fit: Simple heuristics for paired comparison. *Theory and Decision, 52*, 29–71.
- Martignon, L. & Schmitt, M. (1999). Simplicity and robustness of fast and frugal heuristics. *Minds and Machines, 9*, 565–593.
- Mata, R., Helversen, B. von & Rieskamp, J. (2010). Learning to choose: Cognitive aging and strategy selection learning in decision making. *Psychology and Aging, 25*, 299–309.
- Mata, R., Helversen, B. von & Rieskamp, J. (2011). When easy comes hard: The development of adaptive strategy selection. *Child Development, 82*, 687–700.
- Mata, R., Schooler, L. J. & Rieskamp, J. (2007). The aging decision maker: Cognitive aging and the adaptive selection of decision strategies. *Psychology and Aging, 22*, 796–810.
- Mayselless, O. & Kruglanski, A. W. (1987). What makes you so sure? Effects of epistemic motivations on judgmental confidence. *Organizational Behavior and Human Decision Processes, 39*, 162–183.
- Meyer, J., Feinshreiber, L. & Parnet, Y. (2003). Levels of automation in a simulated failure detection task. *IEEE International Conference on Systems, Man and Cybernetics, 3*, 2101–2106.
- Miller, G. (1956). The magical NNN seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review, 63*, 81–97.
- Myung, J. I. (2000). The importance of complexity in model selection. *Journal of Mathematical Psychology, 44*, 190–204.
- Newell, B. R. (2005). Re-visions of rationality? *Trends in Cognitive Sciences, 9*, 11–15.
- Newell, B. R. & Bröder, A. (2008). Cognitive processes, models and metaphors in decision research. *Judgment and Decision Making, 3*, 195–204.

LITERATUR

- Newell, B. R., Rakow, T., Weston, N. J. & Shanks, D. R. (2004). Search strategies in decision making: The success of "success". *Journal of Behavioral Decision Making*, *17*, 117–137.
- Newell, B. R. & Shanks, D. R. (2003). Take the best or look at the rest? Factors influencing "one-reason" decision making. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *29*, 53–65.
- Newell, B. R. & Shanks, D. R. (2007). Perspectives on the "tools" of decision-making. In M. J. Roberts (Hrsg.), *Integrating the mind: Domain general versus domain specific processes in higher cognition* (S. 131–151). Hove: Psychology Press.
- Newell, B. R., Weston, N. J. & Shanks, R. S. (2003). Empirical tests of a fast-and-frugal heuristic: Not everyone "takes-the-best". *Organizational Behavior and Human Decision Processes*, *91*, 82–96.
- Nihm, S. D. (1976). Polynomial law of sensation. *The American Psychologist*, *31*, 808–809.
- Oaksford, M. R. & Chater, N. (1993). Reasoning theories and bounded rationality. In K. I. Manktelow & D. E. Over (Hrsg.), *Rationality: Psychological and philosophical perspectives* (S. 31–60). London: Routledge.
- Oeusoonthornwattana, O. & Shanks, D. R. (2010). I like what I know: Is recognition a non-compensatory determinant of consumer choice? *Judgment and Decision Making*, *5*, 310–325.
- Oppenheimer, D. (2003). Not so fast! (and not so frugal!): Rethinking the recognition heuristic. *Cognition*, *90*, B1–B9.
- Pachur, T., Bröder, A. & Marewski, J. N. (2008). The recognition heuristic in memory-based inference: Is recognition a non-compensatory cue? *Journal of Behavioral Decision Making*, *21*, 183–210.
- Payne, J. W. (1976). Task complexity and contingent processing in decision making: An information search and protocol analysis. *Organizational Behavior and Human Performance*, *16*, 366–387.
- Payne, J. W., Bettman, J. R. & Johnson, E. J. (1988). Adaptive strategy selection in decision making. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *14*, 534–552.
- Payne, J. W., Bettman, J. R. & Johnson, E. J. (1993). *The adaptive decision maker*. Cambridge: University Press.
- Persson, M. & Rieskamp, J. (2009). Inferences from memory: Strategy- and exemplar-based judgment models compared. *Acta Psychologica*, *130*, 25–37.
- Pitt, M. A., Myung, I. J. & Zhang, S. (2002). Toward a method of selecting among computational models of cognition. *Psychological Review*, *109*, 472–490.
- Platzer, C. (2008). *Why simple heuristics not always make us smart: Einflüsse eines konzeptuellen Primings auf die Strategiewahl in einer Multi-Attribut-Entscheidungsaufgabe*. Unpublizierte Diplomarbeit.
- Pohl, R. F. (2006). Empirical tests of the recognition heuristic. *Journal of Behavioral Decision Making*, *19*, 251–271.
- Popper, K. R. (1966). *Logik der Forschung* (2. Aufl.). Tübingen: Mohr.

- Powell, M. & Ansic, D. (1997). Gender differences in risk behaviour in financial decision-making: An experimental analysis. *Journal of Economic Psychology*, 18, 605–628.
- Rakow, T., Newell, B. R., Fayers, K. & Hersby, M. (2005). Evaluating three criteria for establishing cue-search hierarchies in inferential judgment. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, 31, 1088–1104.
- Read, S., Vanman, E. & Miller, L. (1997). Connectionism, parallel constraint satisfaction processes, and gestalt principles: (Re)Introducing cognitive dynamics to social psychology. *Personality and Social Psychology Review*, 1, 26.
- Rieskamp, J. (2006). Perspectives of probabilistic inferences: Reinforcement learning and an adaptive network compared. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 32, 1355–1370.
- Rieskamp, J. (2008). The importance of learning when making inferences. *Judgment and Decision Making*, 3, 261–277.
- Rieskamp, J. & Hoffrage, U. (2008). Inferences under time pressure: how opportunity costs affect strategy selection. *Acta Psychologica*, 127, 258–76.
- Rieskamp, J. & Otto, P. (2006). SSL: A theory of how people learn to select strategies. *Journal of Experimental Psychology: General*, 135, 207–236.
- Rim, S., Uleman, J. S. & Trope, Y. (2009). Spontaneous trait inference and construal level theory: Psychological distance increases nonconscious trait thinking. *Journal of Experimental Social Psychology*, 45, 1088–1097.
- Roberts, S. & Pashler, H. (2000). How persuasive is a good fit? A comment on theory testing. *Psychological Review*, 107, 358–367.
- Rumelhart, D. E., McClelland, J. L. & The PDP Research Group. (1987). *Parallel distributed processing: Explorations in the microstructure of cognition (Volume 1: Foundations)*. Cambridge: MIT Press.
- Sassenberg, K. & Moskowitz, G. B. (2005). Don't stereotype, think different! Overcoming automatic stereotype activation by mindset priming. *Journal of Experimental Social Psychology*, 41, 506–514.
- Sassenberg, K., Moskowitz, G. B., Jacoby, J. & Hansen, N. (2007). The carry-over effect of competition: The impact of competition on prejudice towards uninvolved outgroups. *Journal of Experimental Social Psychology*, 43, 529–538.
- Schlink, S. & Walther, E. (2007). Kurz und gut: Eine deutsche Kurzskaala zur Erfassung des Bedürfnisses nach kognitiver Geschlossenheit. *Zeitschrift für Sozialpsychologie*, 38, 153–161.
- Schooler, L. J. & Hertwig, R. (2005). How forgetting aids heuristic inference. *Psychological Review*, 112, 610 – 628.
- Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics*, 6, 461–464.
- Shah, J. Y., Kruglanski, A. W. & Thompson, E. P. (1998). Membership has its (epistemic) rewards: need for closure effects on in-group bias. *Journal of Personality and Social Psychology*, 75, 383–393.
- Simon, H. (1956). Rational choice and the structure of the environment. *Psychological Review*, 63, 129–138.

LITERATUR

- Simon, H. (1990). Invariants of human behavior. *Annual Review of Psychology*, *41*, 1–19.
- Smith, P. K. & Trope, Y. (2006). You focus on the forest when you're in charge of the trees: Power priming and abstract information processing. *Journal of Personality and Social Psychology*, *90*, 578–596.
- Strull, T. K. & Wyer, R. S. (1979). The role of category accessibility in the interpretation of information about persons: Some determinants and implications. *Journal of Personality and Social Psychology*, *37*, 1660–1672.
- Stanislaw, H. & Todorov, N. (1999). Calculation of signal detection theory measures. *Behavior Research Methods, Instruments, & Computers*, *31*, 137–149.
- Stanovich, K. E. & West, R. F. (2003). Evolutionary versus instrumental goals: How evolutionary psychology misconceives human rationality. In D. E. Over (Hrsg.), *Evolution and the Psychology of Thinking: The Debate* (S. 171–230). Hove: Psychology Press.
- Stapel, D. A. & Koomen, W. (2001). The impact of interpretation versus comparison mindsets on knowledge accessibility effects. *Journal of Experimental Social Psychology*, *37*, 134–149.
- Stapel, D. A. & Suls, J. M. (2007). *Assimilation and contrast in social psychology*. New York: Psychology Press.
- Swets, J. A., Dawes, R. M. & Monahan, J. (2000). Psychological science can improve diagnostic decisions. *Psychological Science in the Public Interest*, *1*, 1–26.
- Tinbergen, N. (1996). The hierarchical organization of nervous mechanisms underlying instinctive behaviour. *Foundations of animal behavior: Classic papers with commentaries*, 406–413.
- Todd, P. M. & Dieckmann, A. (2005). Heuristics for ordering cue search in decision making. *Advances in Neural Information Processing*, *17*, 1393–1400.
- Todd, P. M. & Gigerenzer, G. (2000). Précis of simple heuristics that make us smart. *Behavioral and Brain Sciences*, *23*, 727–780.
- Tversky, A. & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, *185*, 1124–1131.
- Tversky, A. & Kahneman, D. (1981). The framing of decisions and the psychology of choice. *Science*, *211*, 453–458.
- Tversky, A. & Kahneman, D. (1983). Extensional versus intuitive reasoning: The conjunction fallacy in probability judgment. *Psychological Review*, *90*, 293–315.
- Uleman, J. S. & Bargh, J. A. (1989). *Unintended thought*. New York: The Guilford Press.
- Unkelbach, C. (2006). The learned interpretation of cognitive fluency. *Psychological Science*, *17*, 339–345.
- Unkelbach, C. (2007). Reversing the truth effect: Learning the interpretation of processing fluency in judgments of truth. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *33*, 219–230.
- Walther, E., Müller, D. & Schott, O. (2001). Automatisches soziales Verhalten. Wie wirkt sich die Aktivierung der Konzepte Altruismus und Egoismus auf Hilfeleistung aus. *Zeitschrift für Experimentelle Psychologie*, *48*, 248–257.
- Wason, P. C. (1960). On the failure to eliminate hypotheses in a conceptual task. *The Quarterly Journal*

- of *Experimental Psychology*, 12, 129–140.
- Webster, D. (1993). Motivated augmentation and reduction of the overattribution bias. *Journal of Personality and Social Psychology*, 65, 261–271.
- Webster, D. M., Kruglanski, A. W. & Pattison, D. A. (1997). Motivated language use in intergroup contexts: Need-for-closure effects on the linguistic intergroup bias. *Journal of Personality and Social Psychology*, 72, 1122–1131.
- Werth, L. & Förster, J. (2007). Regulatorischer Fokus. *Zeitschrift für Sozialpsychologie*, 38, 33–42.
- Wyer, R. S. & Bargh, J. A. (1997). *The automaticity of everyday life: Advances in social cognition*. New Jersey: Lawrence Erlbaum.
- Zuckerman, M., Koestner, R. & Driver, R. (1981). Beliefs about cues associated with deception. *Journal of Nonverbal Behavior*, 6, 105–114.

Abbildungsverzeichnis

| | | |
|------|--|----|
| 1.1. | Das Brunswiksche Linsenmodell (vgl. Brunswik, 1955, Fig. 8, S. 206). Der Entscheider schließt anhand von Cues auf den Zustand der distalen Variablen. Ökologische Validität und Cue-Nutzung beschreiben den Zusammenhang zwischen der distalen Variablen, den Cues und der zentralen Antwort (resp. Urteil). | 23 |
| 1.2. | Die Oberfläche eines 3 Optionen x 4 Cues Börsenspiels. Ein Proband setzt 2% des möglichen Gewinns für die Konsultation des Cues „Umsatzsteigerung“ für Aktie 1 ein. | 39 |
| 1.3. | Die Oberfläche der Feedback Maske eines drei Optionen Börsenspiels. Angezeigt wird der Gewinn jeder Aktie, die Investition, die ein Proband in einem Durchgang durch die Konsultation von Cues gemacht hat, wie viel Gewinn die Aktie nach Abzug der Investition gemacht hat und der aktuelle Kontostand. | 40 |
| 1.4. | Die Oberfläche der Pferdewette. Ein Entscheider muss in einer simulierten Pferdewette zwischen vier Pferden auf das Siegerpferd tippen. Ein Proband mit einem aktuellen Kontostand von 910 Euro Spielgeld hat drei Hinweisreize aufgedeckt, die nur mit einer bestimmten Validität das Gewinnerpferd anzeigen. | 61 |
| 2.1. | Die Oberfläche des Kriminalfalls. Ein Entscheider muss in einem simulierten Kriminalfall von 3 Verdächtigen den Täter identifizieren. Ein Proband mit einem aktuellen Kontostand von 340 Euro Spielgeld hat 2 Hinweisreize aufgedeckt, die nur mit einer bestimmten Validität den korrekten Täter anzeigen. | 71 |

| | | |
|------|---|-----|
| 2.2. | Die Oberfläche des Bilderrätsels zu zwei Zeitpunkten. Ein Objekt - eine Uhr - befindet sich unter einer Maske aus 100 schwarzen Kästchen. Alle 400 Millisekunden wird 1 Kästchen per Zufall aufgedeckt, was 1 Punkt Spielwährung kostet. Die Probanden sollen erraten, welches Objekt sich unter der Maske befindet, sobald sie sicher genug sind, das Objekt identifiziert zu haben. . . | 73 |
| 3.1. | Aufgabe Manipulation Check <i>Need For Cognitive Closure</i> . Es werden stark vergrößerte Alltagsobjekte präsentiert (a, c). Die Probanden sollen Hypothesen generieren, um welches Objekt es sich handelt. Danach werden die Lösungen zu den Durchgängen präsentiert (b, d). | 116 |
| 4.1. | Gezeigt werden zwei Handys, bei denen alle Elemente kontrolliert wurden. Handy A ist fehlerfrei, Handy B hat vier schwarze fehlerhafte Elemente, die ein Quadrat bilden, und ist somit defekt. | 126 |
| 4.2. | Gezeigt werden zwei Handys, bei denen <i>nicht</i> alle Elemente kontrolliert wurden. Ein schwarzes Element kann bedeuten, dass das Element fehlerhaft ist oder dass das Feld nicht kontrolliert wurde. Handy C ist funktionsfähig, Handy D kann defekt sein (siehe Text). | 127 |
| 4.3. | Die Oberfläche der Kategorisierungsaufgabe. Ein Proband muss entscheiden, ob das Handy auf dem Fließband fehlerhaft ist und entsorgt werden muss oder fehlerfrei ist und verkauft werden kann. | 128 |
| C.1. | Oberfläche der Pferdewette (PHP/HTML Version) in Studie 8 zur Rekognition von Gesichtern (Kapitel 4.1.2, S. 130). | 193 |

Tabellenverzeichnis

| | | |
|------|---|----|
| 1.1. | Validitäten, Diskriminanzraten und Successraten für alle neun Cues der Städtepaarvergleichsaufgabe (vgl. mit Gigerenzer & Goldstein, 1996, Table 1, S. 655). | 26 |
| 1.2. | Beispiel für ein Cue-Muster für vier Cues für den Vergleich zwischen Stuttgart und Nürnberg (vgl. mit Gigerenzer & Goldstein, 1996, Anhang, S. 668). | 28 |
| 1.3. | Erwartete Häufigkeiten der Anwendung einer ORDM und einer MRDM Stoppregel der Cue-Suche bei gegebener Validität des ersten Cues eines Testdurchgangs und der gemessenen DLC Schwelle eines Probanden über das Set von 20 Testdurchgängen der Pferdewette für vier konkurrierende Modelle einer Entscheidung: Das DLC Modell, ein ORDM Modell, ein MRDM Modell und ein Zufallsmodell; vergleiche mit Hausmann und Läge (2008, Figure 2, S. 235). | 64 |
| 2.1. | Kreuztabelle der Modellklassifikation über die 20 Testdurchgänge des DLC der Pferdewette und der Kriminalaufgabe. | 75 |
| 2.2. | Mittelwerte, Streuungen und Korrelationen (Pearson's r) der Sicherheits-Ratings der 13 Entscheidungssituationen mit dem DLC der Pferdewette. | 77 |
| 2.3. | Kreuztabelle der Modellklassifikation über die 20 Testdurchgänge des DLC zu Zeitpunkt t 1 und Zeitpunkt t 2. | 85 |

| | |
|---|-----|
| 2.4. Mittelwert, Standardabweichung, t-Test auf Mittelwertsdifferenzen von Messzeitpunkt 1 zu Messzeitpunkt 2, Effekstärke δ'' der Mittelwertsdifferenzen ¹ für das DLC und die vier Prozessmaße einer Entscheidung im Börsenspiel, die Korrelation aller Variablen über die beiden Messzeitpunkte (in der Diagonalen in Klammern) und die Interkorrelation der Variablen zu Messzeitpunkt 1 (unter der Diagonalen) und zu Messzeitpunkt 2 (über der Diagonalen). | 87 |
| 2.5. Multiple lineare Regression mit dem über die beiden Messzeitpunkte gemittelten DLC als Kriterium und den über die beiden Messzeitpunkte gemittelten vier Prozessmaßen einer Entscheidung als Prädiktoren. | 88 |
| 2.6. Mittelwert des DLC, Streuung des DLC und Anzahl der Probanden, die im Börsenspiel als <i>Take The Best</i> (= ORDM) versus <i>Franklin's Rule</i> und <i>Dawes Rule</i> (= MRDM) Strategienutzer identifiziert wurden; t-Test und Cohen's d zum Mittelwertvergleich des DLC zwischen den beiden Gruppen für den jeweiligen Messzeitpunkt t1 und t2. | 89 |
| 2.7. Kreuztabelle Häufigkeiten der Strategieklassifikation zu Messzeitpunkt 1 und Messzeitpunkt 2 für 21 Probanden, die zu beiden Messzeitpunkten einen Fehlerwert (= strategieinkonsistente Optionswahlen) < .5 haben. . . . | 90 |
| 3.1. Multiple lineare Regression mit dem DLC als Kriterium und den experimentellen Bedingungen als Prädiktoren. | 98 |
| 3.2. Multiple lineare Regression mit der mittleren Anzahl an Gründen für eine Entscheidung als Kriterium und den experimentellen Bedingungen als Prädiktoren. | 99 |
| 3.3. Mittelwert und Streuung für die Bildhaftigkeit, Konkretheit und Bedeutungshaltigkeit gemittelt über die acht Wörter für das jeweilige Set (Baschek et al., 1994). . . . | 103 |
| 3.4. Multiple lineare Regression mit DLC Priming als Kriterium und der experimentellen Bedingung und dem DLC der Testphase als Prädiktoren. | 105 |

4.1. Korrelationen (Pearson's r) zwischen dem DLC und dem *Bedürfnis nach kognitiver Geschlossenheit (16-NCCS)* (Schlink & Walther, 2007), der Differenz des *Promotion* und *Prevention* Fokus (Werth & Förster, 2007), der *Achievement Motives Scale (AMS-R)* (Lang & Fries, 2006) und den *NEO Fünf Faktoren Inventar (NEO-FFI)* der Persönlichkeit (Borkenau & Ostendorf, 1993). 140

A.1. Cuemuster der fünf im Börsenspiel verwendeten Aufgabentypen und die für jeden Aufgabentyp für *Take The Best*, *Dawes Rule*, drei Arten von *Franklin's Rule* und einer *Rate* Strategie vorhergesagten Wahlen. 168

A.2. Validitäten der 10 Cues für die Fälle $DLC > .89$ 172

C.1. Scrambled Sentences der ORDM, der MRDM und der neutralen Bedingung der Studie 4 zum konzeptuellen Priming (vgl. mit Platzer, 2008 und Walther et al., 2001).189

C.2. Mittelwert und Streuung für die Bildhaftigkeit, Konkretheit und Bedeutungshaltigkeit der Begriffe der Mindset Priming Studie (Baschek et al., 1994). 192

A. Formeln

A.1. Bayesianische Strategieklassifikation

Die Bayesianische Strategieklassifikation (Bröder & Schiffer, 2003a) erlaubt Aussagen darüber, wie wahrscheinlich es ist, dass ein Proband eine bestimmte Strategie S_1 gegenüber einer anderen Strategie S_2 unter Berücksichtigung des Entscheidungsverhaltens D des Probanden in einer Entscheidungs-Aufgabe verwendet hat. Das nennt man den a-posteriori Wettquotienten (Bröder, 2005, S. 97) und wird in statistischer Notation als $\frac{p(S_1|D)}{p(S_2|D)}$ dargestellt.

Dabei kann - so lange die dem Entscheidungsverhalten zugrundeliegende Messverteilung spezifiziert ist - ermittelt werden, wie wahrscheinlich es ist, dass die Strategie gegenüber einer anderen Strategie das gemessene Entscheidungsverhalten produziert hat. Das nennt man den Likelihood-Quotienten und wird in statistischer Notation als $\frac{p(D|S_1)}{p(D|S_2)}$ dargestellt.

Die Wahrscheinlichkeit für das Vorliegen einer Strategie nennt man a-priori Wahrscheinlichkeit und wird in statistischer Notation als $p(S_1)$ bzw. $p(S_2)$ dargestellt. Man spricht von „a-priori“, da das die Wahrscheinlichkeit für die Anwendung einer bestimmten Strategie in der Entscheidungsaufgabe *vor* der Berücksichtigung der Evidenz des gemessenen Entscheidungsverhaltens ist.

Der Likelihoodquotient kann über eine Maximum Likelihood Schätzung direkt aus den Daten berechnet werden. Der interessante a-posteriori Wettquotient - in der Strategieklassifikation soll vom Entscheidungsmaß auf die Wahrscheinlichkeit für eine Strategie geschlossen werden - wird über die folgende Formel berechnet (vgl. mit Bröder & Schiffer, 2003a, equation (1), S. 199):

$$\frac{p(S_1|D)}{p(S_2|D)} = \frac{p(S_1)}{p(S_2)} \times \frac{p(D|S_1)}{p(D|S_2)} \quad (\text{A.1})$$

Der Likelihoodquotient wird zur Berechnung des a-posteriori Wett-

A. Formeln

quotienten mit dem Verhältnis der a-priori Wahrscheinlichkeiten für die Strategien gewichtet. Wenn es keine Grund dafür gibt, dass eine Strategie wahrscheinlicher als eine andere Strategie ist - das ist eine gebräuchliche Annahme (Bröder & Schiffer, 2003a, S. 199) -, erhalten alle Strategien die gleiche Wahrscheinlichkeit und der Quotient kürzt sich aus der Gleichung. Der a-posteriori Wettquotient gibt also an, um wieviel mal *wahrscheinlicher* das gemessene Entscheidungsverhalten für S_1 zu S_2 ist.

Um den Likelihoodquotient zu ermitteln, ist es notwendig, für jede Strategie genaue Vorhersagen zum Entscheidungsverhalten, das in einer Entscheidungs-Aufgabe gemessen wird, zu treffen. Im Börsenspiel sind bei drei Optionen und vier binären Cues $\left(\frac{2^4!}{3! \times (2^4 - 3)!}\right) = 560$ verschiedene Cue-Muster möglich (Bröder & Schiffer, 2003a, S. 202). Daraus lassen sich fünf Typen von Aufgaben ableiten.

Tabelle A.1. Cuemuster der fünf im Börsenspiel verwendeten Aufgabentypen und die für jeden Aufgabentyp für *Take The Best*, *Dawes Rule*, drei Arten von *Franklin's Rule* und einer *Rate* Strategie vorhergesagten Wahlen.

| | Itemtyp | | | | | | | | | |
|-------|------------------------|---|-------|---|-------|---|-------|---|-------|---|
| | 1 | | 2 | | 3 | | 4 | | 5 | |
| | A | B | A | B | A | B | A | B | A | B |
| Cue 1 | + | - | + | - | + | - | + | - | + | - |
| Cue 2 | + | + | - | + | - | + | - | - | + | + |
| Cue 3 | + | - | - | + | - | + | - | + | - | + |
| Cue 4 | - | + | - | + | - | - | - | + | - | - |
| | Vorhersage Optionswahl | | | | | | | | | |
| TTB | A | | A | | A | | A | | A | |
| DR | A | | B | | B | | B | | A:B:C | |
| FR1 | A | | B | | A | | A | | A | |
| FR2 | A | | B | | B | | A | | A | |
| FR3 | A | | B | | B | | B | | A | |
| Raten | A:B:C | | A:B:C | | A:B:C | | A:B:C | | A:B:C | |

Anmerkungen: TTB = *Take The Best*, DR = *Dawes Rule* und FR = *Franklin's Rule*; A:B:C = zufällige Wahl zwischen allen Optionen.

In Tabelle A.1 (S. 168) sind beispielhafte Optionen-Cues Muster der

fünf Typen von Entscheidungsdurchgängen und die für jede Strategie abgeleitete Entscheidung für eine der drei Optionen dargestellt (vgl. mit Bröder & Schiffer, 2003a, Table 3, S. 202).

Es werden sechs Strategien miteinander verglichen. Die Cues sind nach der Validität geordnet, so dass Cue 1 der validieste Cue, Cue 2 der zweitvalideste Cue etc. ist. Das Cue-Muster für Option C ist nicht dargestellt, da es in der Aufgabe so gewählt wird, dass unabhängig von den sechs Strategien Option C immer von einer der beiden Optionen dominiert wird. Die Ausnahme bildet die *Raten* Strategie, bei der keine der Optionen dominiert, und *Dawes Rule*, bei der in den Durchgängen des Itemtyps 5 keine Option dominiert.

In Itemtyp 1 führt die Anwendung einer *Take The Best* Strategie beim ersten Cue, der zwischen den Optionen diskriminiert, zu einer Entscheidung für Option A. Die Anwendung von *Dawes Rule* führt für Itemtyp 1 zur Wahl von Option A, da mehr Evidenz für Option A spricht. Bei *Franklin's Rule* wird zwischen drei Unterarten der Strategie unterschieden: In *FR1* kann der valideste Cue nur durch mindestens drei weniger valide Cues übertroffen werden (siehe Itemtyp 1), in *FR2* kann der valideste Cue nur durch mindestens zwei in der Rangordnung direkt folgende weniger valide Cues übertroffen werden (siehe Itemtyp 2) und in *FR3* kann der valideste Cue durch mindestens zwei weniger valide Cues übertroffen werden (siehe Itemtyp 3). Die verschiedenen Arten von *Franklin's Rule* bilden somit eine unterschiedliche absolute Gewichtung der Cues ab. Die Anwendung einer *Raten* „Strategie“ lässt eine zufällige Wahl zwischen den drei Optionen erwarten.

Ein Proband kann demnach unter der Annahme einer Strategie k bei J Itemtypen über ein Set von $\sum_{j=1}^J n_j$ Wahlen $\sum_{j=1}^J n_{jk}$ strategiekonforme Wahlen treffen, $\sum_{j=1}^J n_{j\phi}$ strategieinkonforme Wahlen für die unter allen oben beschriebenen Strategien durch Option A und B dominierten Option C treffen und $\sum_{j=1}^J (n_j - n_{jk} - n_{j\phi})$ Wahlen für die strategieinkonforme Option A oder B treffen.

Die Passung des Vektors der aus einer Strategie k abgeleiteten erwarteten Wahlen und des Vektors der tatsächlichen beobachteten Wahlen eines Probanden kann unter der Verteilungsannahme der Binomialverteilung für Wahlen im Likelihood L_k nach folgender Formel berechnet werden (vgl. mit Bröder & Schiffer, 2003a, equation (4), S. 203):

A. Formeln

$$L_k = p(n_{jk}, n_{j\phi}|k) = \prod_{j=1}^J \left(\frac{n_j!}{n_{jk}! \times n_{j\phi}! \times (n_j - n_{jk} - n_{j\phi})!} \right) \times \\ \times (1 - \epsilon_k - \phi_k)^{n_{jk}} \times \epsilon_k^{(n_j - n_{jk} - n_{j\phi})} \times \phi_k^{n_{j\phi}}. \quad (\text{A.2})$$

Gegeben sind also die Häufigkeiten n_j , n_{jk} und $n_{j\phi}$ für jeden Aufgabentyp j und Strategie k . Geschätzt werden müssen die Parameter ϵ_k , das ist die Fehlerwahrscheinlichkeit für Strategie k die strategieinkonsistente Option A oder B zu wählen, und ϕ_k , das ist die Fehlerwahrscheinlichkeit für Strategie k die von Option A und B dominierte Option C zu wählen. Die Parameter können in einer Maximum Likelihood Algorithmus geschätzt werden, in der der zweidimensionale Parameterraum für die Bestimmung der Parameterausprägungen, die die höchste Likelihood erzielen, abgesucht wird.

Die besten Schätzer für beide Parameter lassen sich jedoch auch über die folgenden Formeln direkt berechnen. Für die Bestimmung des Schätzers $\hat{\epsilon}_k$ ist das (Bröder & Schiffer, 2003a, equation (5), S. 203):

$$\hat{\epsilon}_k = \left[\sum_{j=1}^J (n_j - n_{jk} - n_{j\phi}) \right] \div \left[\sum_{j=1}^J n_j \right]. \quad (\text{A.3})$$

Für die Bestimmung des Schätzers $\hat{\phi}_k$ ist das (Bröder & Schiffer, 2003a, equation (6), S. 203):

$$\hat{\phi}_k = \left(\sum_{j=1}^J n_{j\phi} \right) \div \left(\sum_{j=1}^J n_j \right). \quad (\text{A.4})$$

Für die Raten Strategie ist die Wahrscheinlichkeit für ϵ_k und ϕ_k per Definition bei 1/3 festgelegt: Das ist die Wahrscheinlichkeit für die strategieinkonsistente Wahl bei einer zufälligen Wahl zwischen drei Optionen. Das gilt auch für *Dawes Rule* bei Itemtyp 5.

Es kann somit für jede Strategie die Likelihood berechnet werden. Die Likelihood wird dann im Bayesianischen Informations-Kriterium (Schwarz, 1978), das um die Anzahl der freien Parameter korrigiert, verrechnet (siehe auch Kruschke, 2010). Bei einer Likelihood L für Strategie k , einer Anzahl von N_{obs} beobachteten Entscheidungen und einer Anzahl

von N_p freien Parametern wird das Bayesianische Informationskriterium BIC nach folgender Formel berechnet (vgl. mit Glöckner, 2009, Formel 9, S. 191):

$$BIC = -2 \ln(L_k) + \ln(N_{obs}) \times N_p \quad (A.5)$$

Es werden für alle Strategien im Beispiel - bis auf die Strategie *Raten*, die keine freien Parameter besitzt - mit den beiden Fehlerwahrscheinlichkeiten zwei Parameter geschätzt. Ein Proband wird dann als Nutzer der Strategie klassifiziert, die unter den beobachtetem Entscheidungsverhalten am wahrscheinlichsten ist unter der Annahme, dass die Strategien a-priori gleich wahrscheinlich sind.

A.2. Validitäten erster Cue der DLC Mess- und Testphase

Die Validität des ersten Cues in der Testphase wurde zum Standardverfahren von Hausmann und Läge (2008) in allen Studien - bis auf die chronologisch erste durchgeführte Studie zum Mindset Priming (Kapitel 3.1.2, S. 100) - verändert: Es wurden *nicht* alle Werte zwischen 0 und 1 in .02 Schritten über 51 Durchgänge eingespielt¹, sondern nur Validitäten zwischen .5 und 1 in .01 Schritten präsentiert. Das erlaubt eine höhere Messgenauigkeit im Bereich .5 bis 1, in dem sich, wie die Studie zum Mindset Priming zeigt, alle Probanden befanden.² Die Validitäten der Cues 2 bis 7 wurden genauso wie die Validitäten des ersten Cues gesetzt, das heißt alle Werte zwischen .5 und 1 in .01 Schritten zufällig eingespielt.

Wenn das DLC größer als .79 ist, ist der Maximalwert der 10 Testdurchgänge mit $DLC + .21$ über dem Wert von 1, was inhaltlich nicht möglich ist. Es kann keinen Cue geben, der valider als 1 ist. Daher müssen die Validitäten des ersten Cues der 10 Durchgänge über dem DLC angepasst werden.³

Die Validitäten des ersten Cues der Testdurchgänge - ValTest - für ein DLC zwischen .80 und .90 wurde in allen Studien - bis auf die

¹siehe Kapitel 1.3.2.2 (S. 62)

²Inwieweit eine daraus folgende Einschränkung der Breite des Messbereichs des DLC, da DLC kann dann nur noch Werte zwischen .5 und 1 annehmen, problematisch ist, wird in der Abschluss-Diskussion in Kapitel 5 (S. 141) besprochen.

³Hausmann und Läge (2008) machen keine Angaben zu diesen Spezialfällen.

A. Formeln

Studie zur Rekognition von Gesichtern⁴ (Kapitel 4.1.2, S. 130) - über die folgende Formel berechnet:

$$\text{ValTest}_i = \text{DLC} + ((\text{int}((100 \times (1 - \text{DLC})) \div 2)) \div 100) + v_i$$

$$\vec{v} = (-.05 \ -.04 \ -.03 \ -.02 \ -.01 \ 0 \ .01 \ .02 \ .03 \ .04) \quad (\text{A.6})$$

Anmerkungen: *int* „schneidet“ die Nachkommastellen einer Dezimalzahl ab, das heißt zum Beispiel $\text{int}(5.8) = \text{int}(5.1) = 5$.

Das heißt also zum Beispiel, dass für ein DLC von .80 die Validitäten des ersten Cues in den Testdurchgängen bei $\text{ValTest} = (.85 \ .86 \ .87 \ .88 \ .89 \ .9 \ .91 \ .92 \ .93 \ .94)$ liegen. Für alle Fälle mit einem $\text{DLC} > .90$ sind die Werte des ersten Cues der 10 Durchgänge, die über dem DLC liegen, in Tabelle A.2 (S. 172) aufgelistet.

Tabelle A.2. Validitäten der 10 Cues für die Fälle $\text{DLC} > .89$

| DLC | Validität 1. Cue Testdurchgänge (ValTest) | | | | | | | | | |
|-----|---|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| | C1 | C2 | C3 | C4 | C5 | C6 | C7 | C8 | C9 | C10 |
| .91 | .92 | .93 | .94 | .95 | .95 | .96 | .97 | .98 | .99 | 1 |
| .92 | .93 | .94 | .95 | .95 | .96 | .96 | .97 | .98 | .99 | 1 |
| .93 | .94 | .95 | .95 | .96 | .96 | .97 | .97 | .98 | .99 | 1 |
| .94 | .95 | .95 | .96 | .96 | .97 | .97 | .98 | .98 | .99 | 1 |
| .95 | .95 | .96 | .96 | .97 | .97 | .98 | .98 | .99 | .99 | 1 |
| .96 | .96 | .97 | .97 | .98 | .98 | .98 | .99 | .99 | 1 | 1 |
| .97 | .98 | .98 | .98 | .98 | .99 | .99 | .99 | 1 | 1 | 1 |
| .98 | .99 | .99 | .99 | .99 | .99 | 1 | 1 | 1 | 1 | 1 |
| .99 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |

Anmerkungen: C = Cue.

In der chronologisch ersten durchgeführten Studie zum Mindset Priming (Kapitel 3.1.2, S. 100) wurde - wie in der Originalstudie von Hausmann und Läge (2008) - die Validität des ersten Cues in der Testphase

⁴In dieser Studie wurden alle Validitäten > 1 gleich 1 gesetzt.

A.2. Validitäten erster Cue der DLC Mess- und Testphase

zwischen 0 und 1 in .02 Schritten zufällig eingespielt. Das bedeutet, dass Probanden auch ein $DLC < .21$ haben konnten, was bei einem Minimalwert von $DLC = -.21$ für die Validitäten der Hinweisreize in der Testphase zu Validitäten unter 0 führt, was inhaltlich nicht sinnvoll ist. Ein $DLC < .21$ führte in der ersten Studie zum Mindset Priming zu einem Abbruch der Studie ($n = 0$), da ein $DLC < .25$ auch inhaltlich problematisch ist. Ein $DLC < .25$ bedeutet, dass auch Cue Validitäten, die *unter* der Ratewahrscheinlichkeit von .25 liegen, als einziger Cue in einer Entscheidung akzeptiert werden.

Hier offenbart sich aber auch eine Schwäche des Modells: Die Validität eines Hinweisreizes von zum Beispiel 0 für ein Pferd hat einen hohen Informationswert. Es bedeutet, dass das Pferd des Tipps ganz sicher nicht das Gewinnerpferd sein wird, was die Ratewahrscheinlichkeit von .25 auf .33 erhöht. Aus den Gründen der Messgenauigkeit und aus den genannten konzeptuellen Gründen wurden daher die Validitäten - wie oben beschrieben - in allen anderen Studien verändert.

B. Instruktionen Studien

B.1. Instruktion der Pferdewette

In¹ der nun folgenden Aufgabe schlüpfen Sie in die Rolle des Besitzers eines Wettbüros für Pferderennen.

Das Spannende an Pferderennen für die Zuschauer ist nicht das Rennen und der Sieger allein, sondern das Setzen von Geld auf eines der startenden Pferde. Sie haben einen lukrativen Nebenjob entdeckt, indem Sie die Wetteinsätze von Wettkunden verwalten.

Einzelne Kunden vertrauen Ihnen jeweils 1000 Euro pro Pferderennen an, die Sie jeweils auf eines von mehreren Pferden setzen können.

Wenn das Pferd, auf welches Sie gesetzt haben, das Rennen gewinnt, bekommt der Kunde einen mehrfachen Gewinn zurück und Sie erhalten eine Provision von 300 Euro auf Ihre eigenes Spielkonto ausbezahlt. Wenn Sie auf das falsche Pferd gesetzt haben, gehen sowohl der Kunde als auch Sie leer aus.

Indem Sie jeweils vor den Rennen mit Pferdebesitzern, Jockeys, Rennveranstaltern etc. sprechen, erhalten Sie wertvolle Insiderinformationen, z. B. über die Tagesform einzelner Pferde, des Jockeys, der Anzahl der anspruchsvollen Zuschauer etc., welche die Gewinnwahrscheinlichkeiten zugunsten eines oder mehrerer Pferde verändern.

Die befragten Insider geben Ihnen an, welches der Pferde sie selber als Sieger des aktuellen Rennens sehen. Von jeder Person kennen Sie den Expertisegrad, also die Angabe, wie viele korrekte Vorhersagen sie in den letzten 100 Rennen getroffen hat. Diese Zahl schwankt zwischen 0 bis 100 korrekter Vorhersagen!

Das Einholen von Insiderinformationen ist mit gewissen Kosten verbunden. Das Befragen einer einzelnen Person kostet Sie jeweils 30 Euro und wird Ihnen direkt vom Spielkonto abgezogen. Versuchen Sie, Ihr Spielkonto zu maximieren, indem sie möglichst viele korrekte Vorher-

¹Es folgt die Instruktion, die in den Studien dieser Arbeit verwendet wurde. Sie ist aber bis auf die Währung Schweizer Franken versus Euro identisch zu Hausmann und Läge (2008); vergleiche auch mit Hausmann-Thürig (2004, S. 95).

sagen machen und entsprechend oft die Provision kassieren.

B.2. Studie 1: Instruktion des Börsenspiels

In² der nun folgenden Aufgabe werden Sie sich mit Aktienspekulationen an einer virtuellen Börse beschäftigen. Dabei wird es darauf ankommen, Ihr Kapital möglichst gewinnbringend in Aktien zu investieren.

Diese Börse funktioniert nach ihren eigenen Gesetzen.

Sie bekommen jeweils ein Angebot, das aus drei Aktienpaketen besteht. Ihre Aufgabe wird nun darin bestehen, sich auf der Grundlage verschiedener Informationen, die Ihnen zur Verfügung gestellt werden, für eine dieser Aktien zu entscheiden.

Bei jeder Wahl investieren Sie 1000 Euro Spielwährung in die gewählte Aktie. Je erfolgreicher sich die gewählte Aktie im Vergleich zu den nicht gewählten Aktien entwickelt, desto höher ist natürlich Ihr Gewinn, der Ihrem persönlichen Konto zugeschlagen wird.

Um zwischen den Aktien wählen zu können, werden Ihnen eine Reihe von Informationen zur Verfügung gestellt, die die Erfolgswahrscheinlichkeit der Aktie mitbestimmen. In der REIHENFOLGE IHRER WICHTIGKEIT sind das die AntBegriffen auf folgenden Fragen:

- * Gab es in den letzten Monaten eine UMSATZSTEIGERUNG des Unternehmens?
- * Konnte das Unternehmen eine GEWINNSTEIGERUNG erzielen?
- * Wurde die Ausschüttung der DIVIDENDE pro Aktie erhöht?
- * Hat das Unternehmen im letzten Jahr NEUE STANDORTE eröffnet?

Die Unternehmensmerkmale ermöglichen natürlich - wie bei der echten Börse - keine perfekte Vorhersage des Erfolgs einer Aktie. Sie können aber hilfreich sein, um die Wahrscheinlichkeit des Erfolgs einer Aktie zu ermitteln.

D.h.: Die Aktie eines Unternehmens mit Umsatzsteigerung (bzw. Gewinnsteigerung, Dividendenerhöhung, neu eröffneten Standorten) wird

²Die Instruktion wurde - bis auf nicht erwähnenswerte minimale Veränderungen - aus den Studien zum Börsenspiel von Bröder (2005) übernommen.

B.3. Studie 2: Instruktion des Kriminalfalls

im Durchschnitt wohl erfolgreicher sein als Aktien von Unternehmen, die diese Merkmale nicht aufweisen.

Nun ist es hier wie im richtigen Leben: Man bekommt nichts umsonst!

Die Information über die Unternehmensprofile sind Ihnen nicht frei zugänglich. Sie können sie jedoch dem Börsenmakler Dr. Money in beliebiger Anzahl abkaufen. Dr. Money verlangt pro gekaufter Information eine Provision von 2% Ihres erwirtschafteten Aktiengewinns.

Wenn die gewählte Aktie verliert, will Herr Dr. Money damit natürlich nichts zu tun haben und Sie müssen den Verlust selbst tragen. Wenigstens müssen Sie dann aber nichts für die gekaufte Information abführen.

Sie sind nun darauf vorbereitet Ihr eigenes Kapital von 80.000 Euro Spielgeld in insgesamt 80 Entscheidungen zwischen drei Aktien gewinnbringend einzusetzen.

Zur Erinnerung: Jede gekaufte Information kostet Sie 2% des Aktiengewinns. Um Informationen über bestimmte Aktien einsehen zu können, klicken Sie auf das entsprechende, mit einem Fragezeichen versehene Feld. Es wird daraufhin mit „ja“ oder „nein“ angezeigt, ob das entsprechende Merkmal (Umsatzsteigerung, Gewinnsteigerung, Dividendenerhöhung, Neue Standorte eröffnet) bei diesem Unternehmen vorliegt. Die Merkmale werden Ihnen dabei in der Reihenfolge der Wichtigkeit dargeboten.

Eine einzelne Information ist nur solange sichtbar, wie Sie den Mauszeiger im Informationsfeld belassen, danach erscheint wieder ein Fragezeichen. Pro angeforderter Information sind 2% Ihres Gewinns als Provision an den Makler zu entrichten. Wird dieselbe Information von Ihnen mehrfach angefordert, muss allerdings nur einmal dafür bezahlt werden.

Wenn Sie jetzt keine Fragen mehr haben, geht es mit dem Anklicken des Buttons „Aufgabe beginnen“ los.

Wichtig noch: In jedem Durchgang werden die Karten neu gemischt! Sie können also nicht darauf schließen, dass z.B. Aktie 2 die „bessere“ ist, nur weil sie sich im vorangegangenen Durchgang bewährt hat.

B.3. Studie 2: Instruktion des Kriminalfalls

Die Polizei muss eine Serie von Raubüberfällen aufklären. Zu Ihrer Unterstützung heuert sie das Detektivbüro “die 3 Fragezeichen” an. In der folgenden Aufgabe schlüpfen Sie in die Rolle von Justus Jonas und befinden sich mit ihren Kollegen Bob Andrews und Peter Shaw in

B. Instruktionen Studien

ihrem Detektivbüro, denn es gilt täglich einen neuen Fall aufzuklären.

Aus den gelösten Fällen der Vergangenheit haben Sie noch 500 Euro Belohnung in Ihrer Detektivkasse. Dieses Geld benötigen Sie, um sich Informationen von Experten einzuholen. Wenn Sie den Täter korrekt identifizieren, erhalten Sie eine Belohnung von 400 Euro. Wenn sie einen Verdächtigen beschuldigen, der nicht der Täter ist, entfällt die Belohnung.

Auch hier gilt: Ihr Gewinn wird mit dem Gewinn aus der Pferdewette verrechnet und Ihrem Gesamtkonto für den Kinogutschein gutgeschrieben.

Um Informationen über die Tatverdächtigen zu erhalten, können Sie Experten zu Rate ziehen, wie z.B. Psychologen, die die Aussagen der Tatverdächtigen auf ihre Glaubwürdigkeit hin untersucht haben, Polizisten, die die Spuren am Tatort gesichert haben, etc.

Jeder Experte gibt an, wer seiner Einschätzung nach der Täter ist. Aus Akten kennen Sie die Validität der Einschätzung, d.h. Sie wissen, wieviele korrekte Identifikationen der jeweilige Experten in den letzten 100 Fällen getroffen hat. Die Zahl der korrekten Identifikationen kann also zwischen 0 und 100 liegen.

Das Einholen von Expertenwissen ist mit gewissen Kosten verbunden. Das Befragen eines Experten kostet Sie jeweils 80 Euro und wird Ihnen direkt von der Detektivkasse abgezogen. Versuchen Sie, Ihre Kasse zu maximieren, indem Sie möglichst viele Täter identifizieren und entsprechend oft die Belohnung kassieren.

B.4. Studie 5: Instruktion Need For Cognitive Closure

[Instruktion Comedy Videos oder Instruktion Statistik Videos (s. u.)]

Das Experiment ist zeitlich begrenzt, so dass die Zahl der Videos, die Sie im zweiten Abschnitt der Studie anschauen werden, abhängig von der Geschwindigkeit ist, mit der Sie die Aufgaben des ersten Abschnitts der Studie bearbeiten.

Am unteren Rand des Bildschirms sehen Sie eine Aufgaben-Leiste. Hier werden alle Aufgaben in der Reihenfolge, in der sie in der Studie präsentiert werden, angezeigt. Die rot umrandete Aufgabe ist die Aufgabe, die als nächstes folgt. Die Aufgaben-Leiste soll Ihnen zur Orientierung dienen: Sie sehen welche Aufgabe Sie schon bearbeitet haben und welche Sie noch bearbeiten müssen.

Instruktion Comedy Videos

In Phase 2 der Studie (siehe blauer Bereich) dürfen sie Video-Mitschnitte aus aktuellen Comedy-Sendungen anschauen. Sie dürfen zwischen folgenden Comedy-Videos frei wählen:

- (a) 'Harald Schmidt spricht über Heavy Metal Musik'
- (b) 'Helge Schneider interviewt sich selbst'
- (c) 'Oliver Kalkofe parodiert eine Game Show'
- (d) 'Mario Barth erzählt eine Story über ein Treffen mit seinen Fans'
- (e) 'Heinz Gröning erzählt eine Story über seine Freundin'

Nachdem Sie ein Video angeschaut haben, dürfen Sie beurteilen, wie unterhaltsam Sie den Inhalt des Videos fanden.

Instruktion Statistik Videos

In Phase 2 der Studie (siehe blauer Bereich) dürfen Sie Video-Mitschnitte aus Statistik-Vorlesungen anschauen. Sie dürfen zwischen folgenden Statistik-Videos frei wählen:

- (a) 'Testtheorie: True Score Modell'
- (b) 'Zufallsexperiment'
- (c) 'Einführung in R'
- (d) 'Kausalität: Simpson Paradox'
- (e) 'Skalentypen'

Nachdem Sie ein Video angeschaut haben, wird Ihr Verständnis der statistischen Inhalte mit Fragen getestet.

B.5. Studie 7: Instruktion Kategorisierungsaufgabe

In der folgenden Aufgabe „Handy Firma“ schlüpfen Sie in die Rolle eines Mitarbeiters in der Qualitätssicherung einer Handy Firma. Sie nehmen

B. Instruktionen Studien

in der Handy Firma eine wichtige Position ein: Jedes produzierte Handy wird Ihnen, bevor es zum Verkauf in den Handy Shop ausgeliefert wird, zur Kontrolle vorgelegt. Die Herstellung von Handys ist anspruchsvoll und es ist möglich, dass aufgrund von Herstellungsfehlern oder Materialfehlern defekte Handys produziert werden. Sie wissen aus Erfahrung, dass in Ihrer Firma 50% der Handys fehlerfrei sind, d.h. also 50% der Handys, die bei Ihnen in der Firma produziert werden, sind defekt.

Ihre Aufgabe ist es zu entscheiden, ob ein Handy fehlerfrei ist und zum Verkauf in den Handy Shop geschickt werden soll oder ob ein Handy defekt ist und entsorgt werden soll.

Auf den folgenden Bildschirmen wird Ihnen gezeigt, wie die Handys in dieser Aufgabe auf dem Bildschirm dargestellt werden.

Sie sehen über diesem Text zwei Handys.³ Handy A und Handy B. Ein Handy besteht aus 25 Kästchen. Ein Handy ist defekt, wenn 4 schwarze Kästchen ein 2×2 Quadrat bilden.

Handy A ist demnach fehlerfrei, alle Kästchen von Handy A sind weiß. Handy A kann also zum Verkauf in den Shop geschickt werden. Handy B ist demnach defekt, 4 der 25 Kästchen sind schwarz und in einem 2×2 Quadrat angeordnet. Handy B muss also entsorgt werden.

Da der Handymarkt heiß umkämpft ist und Ihre Abteilung unter Zeitdruck steht, ist es Ihren Kollegen aus der Qualitätskontrolle nicht möglich, alle Kästchen eines Handys auf Fehler zu überprüfen. Ihre Kollegen können immer nur einen Teil der Kästchen eines Handys überprüfen. Sie sehen über diesem Text Handy C und Handy D.

Bei einem weißen Kästchen können Sie sich sicher sein: Das Kästchen wurde überprüft und ist fehlerfrei. Ein schwarzes Kästchen kann zwei Dinge bedeuten:

1. Das schwarze Kästchen wurde überprüft und ist defekt.
2. Das schwarze Kästchen wurde nicht überprüft. Es ist dann nicht sicher, ob das Kästchen fehlerfrei oder defekt ist.

Ein schwarzes Kästchen kann also entweder bedeuten, dass das Kästchen nicht überprüft wurde, oder es kann bedeuten, dass das Kästchen überprüft wurde und defekt ist.

Hier kommen 2 Beispiele dazu⁴: Handy C besteht aus 11 weißen und 14 schwarzen Kästchen. Bei Handy C ist die Entscheidung, ob es sich um ein fehlerfreies oder defektes Handy handelt, einfach zu fällen.

³siehe Abbildung 4.1 (S. 126)

⁴siehe Abbildung 4.2 (S. 127)

B.5. Studie 7: Instruktion Kategorisierungsaufgabe

Selbst wenn alle schwarzen Kästchen defekt wären, bildet keines der schwarzen Kästchen ein 2×2 Quadrat. Noch einmal zur Erinnerung: Ein Handy ist nur dann defekt, wenn 4 schwarze Kästchen ein 2×2 Quadrat bilden. Dieser Fall kann bei Handy C nicht eintreten, das Handy ist also definitiv fehlerfrei und kann zum Verkauf in den Handy-Shop geschickt werden.

Handy D besteht auch aus 11 weißen und 14 schwarzen Kästchen. Diesmal ist die Entscheidung, ob es sich um ein defektes oder fehlerfreies Handy handelt, nicht so leicht zu fällen. Es ist möglich, dass alle schwarzen Kästchen, die ein 2×2 Quadrat bilden, defekt sind (Falls Sie das 2×2 Quadrat aus schwarzen Kästchen in Handy D nicht sehen, können Sie dieses mit dem Drücken des Buttons "rot markieren" hervorheben.). In diesem Fall wäre Handy D defekt. Es ist aber auch möglich, dass alle oder ein Teil der 4 schwarzen Kästchen nicht überprüft wurden und deswegen schwarz sind. In diesem Fall wäre Handy D fehlerfrei. Bei Handy D können Sie nicht definitiv feststellen, ob es sich um ein fehlerfreies oder um ein defektes Handy handelt.

Daher noch einmal zur Erinnerung: Insgesamt sind 50% der Handys fehlerfrei und 50% defekt. Für jedes fehlerfreie Handy, das Sie zum Verkauf in den Handy Shop schicken, erhalten sie 100 Euro. Wenn Sie jedoch ein defektes Handy in den Handy Shop zum Verkauf schicken, müssen Sie 100 Euro Strafe zahlen. Wenn Sie ein fehlerfreies Handy entsorgen, dann verlieren Sie für diese Fehlentscheidung 100 Euro. Wenn Sie ein defektes Handy entsorgen, erhalten Sie 100 Euro für diese richtige Entscheidung. Nach jeder Entscheidung erhalten Sie darüber Rückmeldung, ob Ihre Entscheidung korrekt war.

C. Aufgaben Studien

C.1. Studie 2: Entscheidungsszenarien

Szenario 1: Wetter Ausflug

Sie planen für die kommende Woche einen Ausflug mit Freunden. Sie möchten mit Ihren Freunden eine Radtour unternehmen. Eine Radtour macht aber nur Spass, wenn die Sonne scheint. Sie möchten Zeit und Aufwand in die Planung einer Tour nur stecken, wenn die Chance genügend groß ist, dass die Sonne auch scheint.

Wie sicher muss es **MINDESTENS** sein, dass die Sonne scheint, damit Sie die Radtour planen?

Sie können durch Verschieben des Balkens angeben, wie sicher es für Sie mindestens sein muss, dass die Sonne scheint.

0% bedeutet hierbei, dass es definitiv Regen gibt, 100% bedeutet, dass die Sonne definitiv scheint.

Szenario 2: Wechsel Studienort

Sie entscheiden sich, Ihren Studienort zu wechseln, weil Sie mit den Studienbedingungen Ihrer Universität nicht zufrieden sind. Es ist jedoch nicht sicher, dass ein Studienortwechsel auch eine Verbesserung bedeutet. Ein Studienortwechsel ist jedoch definitiv mit Kosten und Aufwand verbunden, wie z.B. den Kosten für den Umzug.

Wie sicher muss es **MINDESTENS** sein, dass Sie durch einen Wechsel Ihre Studienbedingungen verbessern, damit Sie sich für einen Studienortwechsel entscheiden?

Sie können durch Verschieben des Balkens angeben, wie sicher es für Sie mindestens sein muss, dass Sie durch einen Wechsel Ihre Studienbedingungen verbessern.

0% bedeutet hierbei, dass Sie durch einen Wechsel definitiv keine Verbesserung erzielen, 100% bedeutet, dass Sie durch einen Wechsel definitiv eine Verbesserung erzielen.

Szenario 3: Investition Immobilien

Sie möchten Ihr gespartes Geld in Immobilien anlegen. Es ist jedoch nicht sicher, dass Sie mit dem Kaufen von Immobilien Gewinne machen werden. Es besteht die Möglichkeit, dass Sie das Geld bei einer Fehlinvestition verlieren.

Wie sicher muss es MINDESTENS sein, dass eine Geldanlage in Immobilien Gewinne abwirft, damit Sie sich entscheiden, Ihr Geld in Immobilien anzulegen?

Sie können durch Verschieben des Balkens angeben, wie sicher es für Sie mindestens sein muss, dass Sie mit einer Geldanlage in Immobilien Gewinne machen, damit Sie Ihr Geld in Immobilien investieren.

0% bedeutet hierbei, dass Sie durch eine Geldanlage definitiv Verluste erzielen, 100% bedeutet, dass Sie durch eine Geldanlage definitiv Gewinne erzielen.

Szenario 4: Zuverlässigkeit Ebay

Sie möchten einen gebrauchten Fernseher bei Ebay ersteigern. Bei Ebay können Sie das Bewertungsprofil eines Verkäufers einsehen, d.h. Sie können sehen, wie der Verkäufer für Auktionen in der Vergangenheit von anderen Käufern bewertet wurde. In so eine Bewertung fließen alle möglichen Aspekte des Verkaufs hinein, wie z. B. die Qualität der Ware, die Dauer des Versands etc.

Wie zuverlässig muss ein Verkäufer MINDESTENS sein, dass Sie bei ihm auf einen Fernseher bieten?

Sie können durch Verschieben des Balkens angeben, wie zuverlässig der Verkäufer mindestens sein muss, damit Sie auf seine Ware bieten.

0% bedeutet hierbei, dass der Verkäufer definitiv unzuverlässig ist, 100% bedeutet, dass der Verkäufer definitiv zuverlässig ist.

Szenario 5: Wirkung Therapie

Sie bekommen von Ihrem Arzt eine Therapie empfohlen, die gegen Ihre Heuschnupfenallergie helfen soll. Die Therapie schlägt aber nicht bei jedem Patienten an und die Therapie ist mit möglichen Nebenwirkungen wie Kopfschmerzen verbunden.

Wie sicher muss es MINDESTENS sein, dass die Therapie Ihnen hilft, damit Sie die Therapie machen?

Sie können durch Verschieben des Balkens angeben, wie sicher es für Sie mindestens sein muss, dass Ihnen die Therapie hilft, damit Sie die Therapie machen.

0% bedeutet hierbei, dass Ihnen die Therapie definitiv nicht hilft, 100% bedeutet, dass Ihnen die Therapie definitiv hilft.

Szenario 6: Wirkung Impfung

Seit einigen Jahren bekommen Sie im Winter, pünktlich zu Weihnachten, eine schwere Grippe. Eine Grippeimpfung, welche Ihnen der Arzt empfohlen hat, könnte die Wahrscheinlichkeit senken, dass Sie auch dieses Jahr betroffen sind. Jedoch ist die Impfung mit Kosten und dem Restrisiko, dass die Grippe trotz Impfung ausbricht, verbunden.

Wie zuverlässig muss die Impfung MINDESTENS sein, dass Sie sie wahrnehmen?

Sie können durch Verschieben des Balkens angeben, wie zuverlässig die Impfung mindestens sein muss, damit Sie sich impfen lassen.

0% bedeutet hierbei, dass die Impfung definitiv unzuverlässig ist, 100% bedeutet, dass die Impfung definitiv zuverlässig ist.

Szenario 7: Wechsel Job

Sie arbeiten als Angestellte/r in einer Firma. Sie machen Ihren Job gut, fühlen sich aber seit einiger Zeit bezüglich Beförderungen übergangen. Eine konkurrierende Firma bietet Ihnen eine attraktive Stelle an, die mit einer minimalen Lohnerhöhung verbunden wäre. Daraufhin wird Ihnen auch in Ihrer eigenen Firma eine hohe Position in Aussicht gestellt, welche Sie evtl. nächstes Jahr bekleiden könnten und welche mit einer wesentlichen Lohnerhöhung verbunden wäre.

Wie sicher muss Ihnen die Stelle in Ihrer Firma MINDESTENS in Aussicht gestellt werden, dass Sie die Stelle in der konkurrierenden Firma unwiderruflich absagen?

Sie können durch Verschieben des Balkens angeben, wie sicher Ihnen der Job in Ihrer Firma mindestens zugesagt werden muss, damit Sie das andere Angebot ablehnen.

0% bedeutet hierbei, dass die Zusage definitiv unsicher ist, 100% bedeutet, dass die Zusage definitiv sicher ist.

Szenario 8: Treffen Traumpartner

Sie haben Ihren möglichen Traumpartner in einer Online-Singlebörse kennengelernt und sind über sein / ihr Profil hin und weg. Sie haben sich bis jetzt nur telefonisch und über Kurznachrichten unterhalten und dabei festgestellt, dass Ihre Interessen stark einander gleichen.

Wie sicher muss es MINDESTENS sein, dass die Person zu einem Treffen auch kommt, damit Sie sich mit ihr verabreden?

Sie können durch Verschieben des Balkens angeben, wie sicher es für Sie mindestens sein muss, dass diese Person auch wirklich zu der Verabredung kommt.

0% bedeutet hierbei, dass Sie sich sicher sind, dass die Person nicht kommen wird, 100% bedeutet, dass die Person definitiv kommen wird.

Szenario 9: Organisation Party

Sie haben kommende Woche Geburtstag und überlegen derzeit, ob Sie Ihren Geburtstag feiern oder nicht, da Sie knapp bei Kasse sind und nicht wissen, ob alle Ihre guten Freunde an diesem Tag auch Zeit hätten auf Ihre Party zu kommen und sich der finanzielle Einsatz demnach rentieren würde.

Wie sicher muss es MINDESTENS sein, dass Ihre Freunde auch wirklich kommen, dass Sie sich dafür entscheiden, Ihren Geburtstag zu feiern?

Sie können durch Verschieben des Balkens angeben, wie sicher es für Sie mindestens sein muss, dass Ihre Freunde kommen, dass sich für Sie die Investition in Getränke etc. rentiert.

0% bedeutet hierbei, dass Ihre Freunde definitiv keine Zeit haben werden, 100% bedeutet, dass Ihr Freunde definitiv kommen werden.

Szenario 10: Wohnung Partner

Sie haben sich vorgenommen, in den nächsten Wochen mit Ihrem Partner in eine gemeinsame Wohnung zu ziehen. Es ist aber nicht sicher, dass dieser Umzug ohne Konflikte möglich ist.

Wie sicher muss es MINDESTENS sein, dass es nicht zu Konflikten kommt, damit Sie zusammenziehen?

Sie können durch Verschieben des Balkens angeben, wie sicher es für Sie mindestens sein muss, dass Sie ohne Konflikte zusammenziehen.

0% bedeutet hierbei, dass es definitiv Konflikte gibt, 100% bedeutet, dass es definitiv konfliktfrei abläuft.

Szenario 11: Unruhen Reiseland

Sie bekommen von einer Freundin eine Reise empfohlen. Sie haben jedoch in den Nachrichten gehört, dass es dort Unruhen gibt, jedoch nicht überall im Land.

Wie sicher muss es MINDESTENS sein, dass die Unruhen vor Reiseantritt beendet sind, damit Sie die Reise antreten?

Sie können durch Verschieben des Balkens angeben, wie sicher es für Sie mindestens sein muss, dass die Unruhen vor Reiseantritt vorbei sind, damit Sie reisen.

0% bedeutet hierbei, dass Unruhen definitiv bestehen, 100% bedeutet, dass die Unruhen definitiv vorbei sind.

Szenario 12: Energiesparen Hausbau

Sie und ihr/e Ehemann/Ehefrau planen ein Haus zu bauen. Dabei bietet ihr Bauunternehmer eine Bauweise an, bei der Styroporplatten zwischen der Außenmauer und den Innenputz gelegt werden. Natürlich ist diese Bauweise mit höheren Kosten und Aufwand verbunden.

Es ist allerdings nicht sicher, ob Sie dadurch viel Energie sparen können werden.

Wie sicher muss es MINDESTENS sein, dass Sie durch die spezielle Bauweise Energie einsparen können, damit Sie sich für die Bauweise entscheiden?

0% bedeutet hierbei, dass definitiv keine Energie eingespart wird, 100% bedeutet, dass definitiv Energie eingespart wird.

Szenario 13: Funktionalität Prothese

Ihr Orthopäde empfiehlt ihnen anstatt einer normalen Hüftprothese eine neuere Prothese, die erst seit kurzem auf dem Markt ist und aufgrund ihrer neuartigen Verarbeitung dem Patienten eine bessere Beweglichkeit ermöglichen soll und sehr langlebig ist. Allerdings ist die Prothese auch um einiges teurer.

Wie sicher muss es MINDESTENS sein, dass die neue Prothese Ihnen eine bessere Beweglichkeit ermöglicht, damit Sie sich für sie entscheiden?

0% bedeutet hierbei, dass Sie definitiv keine Verbesserung haben, 100% bedeutet, dass Sie definitiv eine Verbesserung in der Bewegungsfähigkeit haben.

C.2. Studie 3: Sätze Konzeptuelles Priming

[siehe folgende Seite]

Tabelle C.1. Scrambled Sentences der ORDM, der MRDM und der neutralen Bedingung der Studie 4 zum konzeptuellen Priming (vgl. mit Platzter, 2008 und Walther et al., 2001).

| Satz | ORDM Bedingung | MRDM Bedingung | Neutrale Bedingung |
|------|--|----------------------------|--------------------|
| 1* | Reinigungskräfte, arbeiten, sollten, Sekretärinnen, <i>flott</i> | gründlich | umsichtig |
| 2 | Vögel, westwärts, Wolken, die, ziehen | - | - |
| 3* | Peter, <i>Risiken</i> , muss, <i>eingehen</i> , man | ausführlich, recherchieren | Kompromisse |
| 4* | handeln, Menschen, <i>intuitive</i> , sind, klüger | besonnene | höfliche |
| 5 | erforderten, Umstände, verlangten, Maßnahmen, die | - | - |
| 6* | Bürger, Wähler, braucht, <i>schnellentschlossene</i> , Deutschland | informierte | selbstbestimmte |
| 7 | abends, leer, Straßen, beleuchtet, sind | - | - |
| 8* | sind, überwiegend, Gedanken, <i>spontan</i> , häufig | komplex | abstrakt |
| 9 | Kaffee, ist, der, stark, Gewichtheber | - | - |
| 10* | <i>wenige</i> , berücksichtigt, wurden, Indizien, gesammelt | viele | interessante |
| 11* | erzieherische, erfolgten, therapeutische, <i>rasch</i> , Maßnahmen | bedachtsam | heute |
| 12 | Wasser, das, Becken, sauber, ist | - | - |
| 13* | <i>sparsam</i> , ist, vorzugehen, war, vorteilhaft | langsam | ehrllich |
| 14 | weiß, das, ist, Haus, Auto | - | - |
| 15* | <i>günstige</i> , verfolgt, er, verwendet, Strategien | teure | einleuchtende |
| 16* | Geschick, <i>einfache</i> , Einfallsreichtum, erfordern, Lösungen | genaue | nützliche |
| 17* | leisteten, <i>schnelle</i> , verlangten, Berichterstattung, Journalisten | umfassende | gute |
| 18 | Wände, färben, sie, Felle, die | - | - |
| 19* | <i>unkompliziert</i> , Verfahrensweise, die, war, erschien | aufwändig | angemessen |
| 20 | gestern, Montag, wieder, Meditationskurs, war | - | - |

Anmerkungen: Kursive Wörter in der Spalte *ORDM Bedingung* werden in der MRDM Bedingung bzw. neutralen Bedingung der Studie durch die Wörter in der Spalte *MRDM Bedingung* bzw. *Neutrale Bedingung* ersetzt; * = Priming Sätze / Rest = Distraktoren.

C.3. Studie 3: Szenarien Konzeptuelles Priming

Im Folgenden¹ werden wir Ihnen kurze Szenarien vorstellen, in denen Personen Entscheidungen getroffen haben. Lesen Sie diese bitte aufmerksam durch und notieren Sie im Anschluss daran einen möglichen Grund/mögliche Gründe, der/die zu dieser Entscheidung geführt haben könnten.

Szenario 1

Dr. Peters ist Internist in einem Krankenhaus. Da seine beiden Kolleginnen krank sind, hat er heute alleine Nachtdienst in der Notaufnahme. Um 1 Uhr werden fast zeitgleich zwei Patienten in die Notaufnahme eingeliefert. Einer der Patienten (A) blutet stark und scheint unter Schock zu stehen, der andere (B) weist keine Verletzungen auf, hat aber das Bewusstsein verloren. Dr. Peters beschließt Patient B als ersten zu behandeln.

Szenario 2

Frau Obermaier ist Lehrerin an einer Grundschule. Heute schreibt sie in Klasse 3b eine Mathematik Klausur. Plötzlich fällt ihr auf, dass Anne immer wieder in ihrem Schreibmännchen wühlt, dann innehält und weiter schreibt. Frau Obermaier hat den Verdacht, dass Anne einen Spickzettel versteckt hat. Frau Obermaier entschließt sich dazu Anne nicht sofort, sondern erst nach der Schulstunde darauf anzusprechen.

Szenario 3

Der Sozialarbeiter Herr Herbold ist Leiter des psychosozialen Dienstes in Frankfurt. Gegen Ende des Jahres werden sowohl Geldmittel als auch Personalressourcen wie immer knapper. Beim Durchforsten der Akten stellt er jedoch fest, dass vor kurzem noch Spendengelder eingegangen sind, die er in seinen Planungen nicht berücksichtigt hatte. Seit einiger Zeit sind zwei Projekte angedacht, die jedoch nie realisiert werden konnten: bei Projekt A geht es darum, eine Selbsthilfegruppe für alleinerziehende Mütter/Väter ins Leben zu rufen, in Projekt B sollen Maßnahmen zur sinnvollen Freizeitbeschäftigung für Jugendliche

¹Die Szenarien stammen aus der Validierungsstudie von Platzer (2008)

C.4. Studie 4: Verwendete Begriffe Mindset Priming

initiiert werden. Da die Gelder nur für die Realisierung eines Projekts reichen, entscheidet sich Herr Herbold für Projekt B.

C.4. Studie 4: Verwendete Begriffe Mindset Priming

[siehe folgende Seite]

C. Aufgaben Studien

Tabelle C.2. Mittelwert und Streuung für die Bildhaftigkeit, Konkretheit und Bedeutungshaltigkeit der Begriffe der Mindset Priming Studie (Baschek et al., 1994).

| Set Begriff | Bildhaftigkeit | | Konkretheit | | Bedeutungs- haltigkeit | |
|--------------------------|----------------|-------|-------------|-------|---------------------------|-------|
| | M | SD | M | SD | M | SD |
| 1 Meer | 17.60 | 5.07 | 17.20 | 4.40 | 14.07 | 5.33 |
| Garten | 15.20 | 5.60 | 15.33 | 6.33 | 12.87 | 6.40 |
| Winter | 16.53 | 5.80 | 11.53 | 10.47 | 12.60 | 7.67 |
| Möbel | 16.27 | 5.07 | 13.53 | 6.80 | 12.53 | 6.87 |
| Familie | 15.73 | 6.07 | 3.87 | 12.13 | 12.07 | 7.40 |
| Wald | 17.87 | 4.20 | 16.33 | 5.53 | 12.07 | 5.53 |
| Auto | 17.60 | 3.80 | 15.33 | 6.13 | 12.07 | 6.67 |
| Großstadt | 16.53 | 5.47 | 10.73 | 7.80 | 11.33 | 8.40 |
| 2 Erdbeere | 18.40 | 2.93 | 19.00 | 2.93 | 6.67 | 8.87 |
| Klavier | 15.47 | 7.60 | 17.20 | 5.00 | 6.67 | 8.27 |
| Sessel | 16.80 | 5.47 | 17.93 | 3.60 | 6.67 | 7.87 |
| Sturm | 14.67 | 6.93 | 13.73 | 6.47 | 6.93 | 6.00 |
| Umweltver- schmutzung | 14.13 | 8.67 | 10.33 | 12.13 | 6.40 | 10.60 |
| Gefängnis | 16.00 | 5.47 | 15.53 | 5.67 | 6.40 | 7.07 |
| Alkohol | 12.27 | 6.87 | 15.73 | 6.60 | 4.20 | 8.27 |
| Gras | 16.00 | 5.07 | 17.00 | 5.53 | 6.67 | 8.87 |
| 3 Standpunkt | 0.00 | 12.93 | -2.20 | 10.33 | -7.67 | 9.00 |
| Augenblick | -1.60 | 10.93 | -3.67 | 11.80 | -7.67 | 7.47 |
| Einbildung | -2.13 | 12.40 | -8.27 | 10.40 | -7.67 | 8.60 |
| Anregung | -5.07 | 11.73 | -4.47 | 8.80 | -7.67 | 7.93 |
| Wirkung | 0.00 | 10.20 | -2.40 | 11.53 | -7.67 | 8.13 |
| Vorfall | -2.13 | 12.73 | -1.20 | 10.87 | -7.67 | 6.87 |
| Verrat | -1.60 | 11.60 | -2.20 | 9.20 | -7.67 | 7.53 |
| Folge | 1.07 | 10.67 | -2.20 | 10.20 | -7.67 | 7.33 |
| 4 Zusatz | -4.80 | 10.07 | -3.47 | 11.33 | -12.33 | 7.07 |
| Ehrenhaftigkeit | -4.13 | 12.73 | -10.53 | 8.53 | -12.40 | 6.60 |
| Überschneidung | -2.67 | 10.20 | -1.40 | 11.53 | -12.60 | 5.33 |
| Fortdauer | -6.67 | 9.20 | -6.47 | 7.73 | -10.40 | 8.13 |
| Einhelligkeit | -10.40 | 10.93 | -10.33 | 9.00 | -13.33 | 6.40 |
| Vorbedacht | 6.40 | 12.40 | -8.67 | 9.93 | -13.87 | 6.53 |
| Hinblick | -9.60 | 8.20 | -6.27 | 9.73 | -14.60 | 6.93 |
| Inbegriff | -12.27 | 9.73 | -8.87 | 9.53 | -15.33 | 6.87 |

C.5. Studie 8: Oberfläche Pferdewette

| <p>Es sind 4 Pferde am Start</p> <table border="1"> <thead> <tr> <th>Pferd A</th> <th>Pferd B</th> <th>Pferd C</th> <th>Pferd D</th> </tr> </thead> <tbody> <tr> <td>Sol</td> <td>Hachman</td> <td>Nanna</td> <td>Harmonia</td> </tr> </tbody> </table> | Pferd A | Pferd B | Pferd C | Pferd D | Sol | Hachman | Nanna | Harmonia | <p>Aufgabe Nr. 1</p> <p>Sie setzen auf Pferd <input type="text" value=""/></p> <p><input type="button" value="OK"/></p> | | | | | | | | | | | | | | | | |
|--|----------|--|--|-----------|-----|---------|-----------|----------|---|-----------|--|--|-----------|--|--|-----------|--|--|-----------|--|--|-----------|--|--|--|
| Pferd A | Pferd B | Pferd C | Pferd D | | | | | | | | | | | | | | | | | | | | | | |
| Sol | Hachman | Nanna | Harmonia | | | | | | | | | | | | | | | | | | | | | | |
| <table border="1"> <thead> <tr> <th>Hinweise</th> <th>Tipp</th> <th>Anzahl korrekter Vorhersagen in den letzten 100 Pferderennen</th> </tr> </thead> <tbody> <tr> <td>Hinweis 1</td> <td>C</td> <td>86</td> </tr> <tr> <td>Hinweis 2</td> <td>C</td> <td>87</td> </tr> <tr> <td>Hinweis 3</td> <td></td> <td></td> </tr> <tr> <td>Hinweis 4</td> <td></td> <td></td> </tr> <tr> <td>Hinweis 5</td> <td></td> <td></td> </tr> <tr> <td>Hinweis 6</td> <td></td> <td></td> </tr> <tr> <td>Hinweis 7</td> <td></td> <td></td> </tr> </tbody> </table> | Hinweise | Tipp | Anzahl korrekter Vorhersagen in den letzten 100 Pferderennen | Hinweis 1 | C | 86 | Hinweis 2 | C | 87 | Hinweis 3 | | | Hinweis 4 | | | Hinweis 5 | | | Hinweis 6 | | | Hinweis 7 | | | <p>Das Aufdecken eines einzelnen Hinweises kostet 30 Euro</p> <p><input type="button" value="nächster Hinweis"/></p> <p>Ihr aktueller Kontostand beträgt 940 Euro.</p> <p>Es muss mindestens ein Hinweis aufgedeckt werden!</p> <p>Ihre Provision bei korrekter Vorhersage beträgt 300 Euro.</p> |
| Hinweise | Tipp | Anzahl korrekter Vorhersagen in den letzten 100 Pferderennen | | | | | | | | | | | | | | | | | | | | | | | |
| Hinweis 1 | C | 86 | | | | | | | | | | | | | | | | | | | | | | | |
| Hinweis 2 | C | 87 | | | | | | | | | | | | | | | | | | | | | | | |
| Hinweis 3 | | | | | | | | | | | | | | | | | | | | | | | | | |
| Hinweis 4 | | | | | | | | | | | | | | | | | | | | | | | | | |
| Hinweis 5 | | | | | | | | | | | | | | | | | | | | | | | | | |
| Hinweis 6 | | | | | | | | | | | | | | | | | | | | | | | | | |
| Hinweis 7 | | | | | | | | | | | | | | | | | | | | | | | | | |

Abbildung C.1. Oberfläche der Pferdewette (PHP/HTML Version) in Studie 8 zur Rekognition von Gesichtern (Kapitel 4.1.2, S. 130).